ISSN 1881-6436

Discussion Paper Series

No. 17-01

ベヴァリッジ『自由社会における完全雇用』のケインズ的要素
～テキストマイニングを加味した量的・質的分析～

小峯敦・下平裕之

2017 年 9 月

〒612-8577 京都市伏見区深草塚本 67
龍谷大学経済学部

タイトル

ベヴァリッジ『自由社会における完全雇用』のケインズ的要素
〜テキストマイニングを加味した量的・質的分析〜

著　者

小峯敦[1]・下平裕之[2]

概　要

本稿は、経済理論がどのように受容され、経済政策の形成に関与するかという大枠の研究プロジェクトに属する。具体的には、ベヴァリッジ『自由社会における完全雇用』(1944)にケインズ『一般理論』(1936)がどの程度、影響しているかを、テキストマイニングという量的分析も加味して考察した。その結果、失業の本質論（貨幣経済論）は脱落し、有効需要の不足という一点に注目が集まっている（問題の単純化）ものの、ベヴァリッジ独自の社会観と処方箋も保持している（独自性の維持）ことも確かめられた。また、国際関係の視点が強いベヴァリッジ、利子率を中心とした貨幣的概念を重視するケインズ、という新たな発見もあった。このように、従来の経済学史の方法論とテキストマイニングを適切に接合することによって、新たな知見を発掘できる可能性がある。

キーワード

テキストマイニング、質的分析、経済学史の方法論、頻度と共起、複合語（専門用語）、完全雇用、戦後計画

JEL classification
B22, B25, B31, B41, B52, C18

[1] 龍谷大学経済学部（〒612-8577 京都市伏見区深草塚本町 67 ；075-645-1111）
komine@econ.ryukoku.ac.jp
[2] 山形大学人文学部（〒990-8560　山形市小白川町一丁目 4-12； 023-628-4203）
shimo@human.kj.yamagata-u.ac.jp

2017.9.12

ベヴァリッジ『自由社会における完全雇用』のケインズ的要素
～テキストマイニングを加味した量的・質的分析～

小峯敦・下平裕之

version 2.2

第１節　序論～理論と政策を結ぶもの

本節では、大枠としての《研究プロジェクト》における本研究の位置づけを述べた後、本稿での仮説を提示する。

1-1　研究プロジェクトの概要

本稿は科研費プロジェクト「経済理論の大衆化から経済政策の形成へ：テキストマイニングを応用した実証研究」[1]（2015-2019）の一環である。さらに本稿は「経済学の制度化論」[2]および「経済政策思想史」[3]という分野に属し、その中でも「政策におけるケインズ革命」[4]（20 世紀前半のイギリスが中心）という論題に注目している。そして「経済問題と認知される領域が広範であった 20 世紀前半の考察には、社会保障・完全雇用の考察を一対として捉える」（小峯 2014: 38-39）という分析視角を持っている。

まず、このプロジェクトの概要を述べる前に、その前身となる研究計画について、説明しておこう。その前身は「経済思想の受容・浸透過程に関する実証研究」[5]（2010-2014）である。このプロジェクトは経済学的思考が一般大衆である非専門家にどのように伝播するか、その過程を

(a)200 年という大きな射程で、

(b)イギリスに焦点を絞り、

(c)質的および量的に特定化・類型化する、

という試みであった。ただし、従来の「経済学の制度化論」や「経済政策思想史」という枠組みの先行研究においても、「経済学の通俗化」、「大衆を経由した政策形成論」については考慮が不十分であった。このプロジェクトは革新的なアイデアが人々に到達する最終段階まで視野を広げている。

経済思想の浸透過程を測る具体的な一次的接近方法として、次のような視角が念頭にあった。その時代を象徴する支配的な経済学の書物（例：アダム・スミス『国富論』、カール・マルクス『資本論』、ジョセフ・シュンペーター『経

---

[1]「経済理論の大衆化から経済政策の形成へ：テキストマイニングを応用した実証研究」（基盤研究 B、研究代表者：下平裕之、課題番号 15H03330、2015-2019）。第 59 回経済思想研究会（2017.8.6；於：東北学院大学）におけるコメントに感謝する。

[2] Coats (1981) (1993)や池尾（2006）が代表的な論考である。

[3] Furner and Supple eds. (1990)、西沢・服部・栗田編（1999）を嚆矢とする。

[4] この研究動向については、小峯（2014: 38-39）で詳述したので、ここでは割愛する。

[5]「経済思想の受容・浸透過程に関する実証研究：人々は経済学をどのように受け入れたか」（基盤研究 B、研究代表者：下平裕之、課題番号 22330064、2010-2014）。

済発展の理論』、J. M. ケインズ『雇用・利子および貨幣の一般理論』など）に込められた思想が、多くの人々の目に触れていた評論・入門書・教科書・解説書・通俗書・詩・寓話・小説・パンフレット・ビラなどを通じて、どのように人々に受け入れられていったのだろうか。その過程で、どのような《変形》・《変換》・《通俗化》が行われたのだろうか。

このプロジェクトは伝統的な経済学史の方法を基調としつつ、新機軸として、テキストマイニングという方法を試行錯誤しながら導入した。外部の知見や他分野の先行者を招聘し、あるいは出向いて、新しい方法の修得に努めたのである。従来の経済学史や経済思想史が用いてきた質的分析に加えて、以下で示されるように、デジタル化されたテキストを用いて、専門用語の出現頻度に基づいた量的解析（記号のグループ化：クラスター分析・主成分分析など）を試みた。このような手続きを経て、このプロジェクトでは、有力な経済学がどのように変形され、受容され、最終的には経済政策や通念として社会を動かす力になったのか／あるいはならなかったのか、このリンクを動態的に描き出すことを目指したのである。この成果の一端は、下平・小峯・松山（2012）、下平・福田（2014）、古谷（2014）、Matsuyama (2016)という形で与えられている。

この５年間で得られた知見を前提に、次の５年間のプロジェクトが次のように定まった（図 1-1 を参照）。

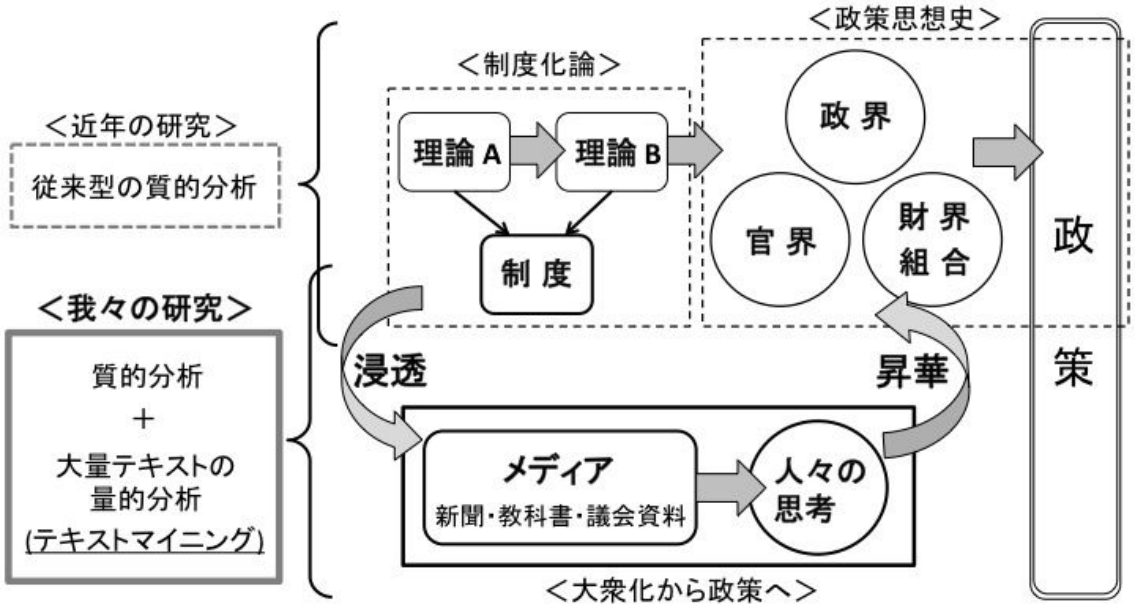


図 1-1　研究プロジェクトの全容

本研究は経済理論が経済政策に影響を及ぼすプロセス[6]について、二重に新しい分析アプローチによる実証を目的とする。（1）経済理論から経済政策への直接的影響という従来の手法とは別に、経済理論が社会に受容されて世論の一部を形成し、その世論が立法過程に影響を及ぼすという媒介プロセスを辿ることを試みる。この視角によって、政策形成を左右する一般の人々の思考や世論を視野に入れることができる。（2）この分析には大量の活字データを扱わざるを得ないので、従来の質的分析とテキストマイニングとを統合した分析方法を開発する。この方法によって、対象テキストの選択が限定的で恣意的になりがちだった従来の方策を乗り越える。

このプロセス設定と分析手法は、議会制の整った 18 世紀後半以降のイギリスにおいて特に有効となる。本研究は 18 世紀後半から 20 世紀前半のイギリスに対象を絞り、東北地方で 15 年以上チームを組んでいる若手・中堅の研究者の協働によって、経済学の影響力を実証する。

1-2　本稿の仮説

以下では、この研究プロジェクトの一環として、１つの仮説を検証することを目的としよう。20 世紀において人々に最大の衝撃を与えた経済書、ケインズ『一般理論』（1936）の影響について、下平・小峯・松山（2012）では一般の読者への理解力という意味で、その書評をテキストマイニングによって分析した。新たな知見としては、書評の発表媒体（新聞・一般雑誌・学術雑誌）ごとに、使用される用語に明らかな差が可視化できることであった。

本稿では、政策担当者への浸透力という意味で、『一般理論』が同時代の顕著な有名人にどのような影響を与えたかを分析する。具体的には、完全雇用政策をイギリス政府に認めさせる文脈で、ケインズと並んで大きな影響力を持ったとされるベヴァリッジ『自由社会における完全雇用』（1944）に、ケインズ的要素がどの程度見られるかを、テキストマイニングによって分析する（以下、『自由社会』と略記する）。

本稿における仮説は、２つの要素から構成される。

---

[6] 詳しくは小峯（2014: 37-40）を参照。Fujita (2014: 338-341)は政策の転換が起こるための条件――危機、責任ある政党、良いアイデア、幸運――がスウェーデンで揃っていたかどうかを議論している。

仮説：ケインズ『一般理論』の政策への影響を考える時、政策決定者や一般読者を想定しているベヴァリッジ『自由社会』では、

(i)失業の本質（貨幣経済）に注目するよりは、有効需要の不足という表面上の原因に着目しつつ（問題の単純化）、

(ii)ベヴァリッジ自身の社会観と処方箋を依然として保持している（独自性の維持）、

という２つの傾向がある。

以下でこの仮説を検証していくが、テキストマイニングという新しい手法を用いるために、第２節では、社会科学の方法論を意識する形で、テキストマイニングと経済学史の接続について、一般論を展開する。第３節では、テキストマイニングに必須な前処理と若干の分析について、具体的な過程を記述する。第４節では、『一般理論』との対比を念頭に、『自由社会』を様々な角度から解析する。第５節では以上の議論をまとめ、今後の課題を付しつつ、結論を述べる。

第２節　テキストマイニングと経済学史の方法

この節[7]では、伝統的な経済学史・経済思想史の方法（主として定性的分析）と比較する形で、テキストマイニングを簡単に説明し、さらにこの方法を経済学史研究に導入する意義・限界をまとめておこう。

2-1　経済学史の方法

経済学史[8]History of Economic Thought という研究は、時代を画す理論そのものの内的論理・整合性を吟味する理論史 economics in text、その理論が生まれた社会的背景を問う思想史 economics in context という２つに大別される。この両者はそれぞれ、現在および過去に重きを置いた経済思想の再構成であり、「合理的再構成」および「歴史的再構成」と呼ぶことができるだろう[9]。ここに

---

[7] 本節の 2-1, 2-2 の一部は、下平・小峯・松山（2012）の 1-3 を改訂したものである。

[8] マルクス主義の影響を受けた「経済学説史」という表記法――この場合、「理論史」を強く指向している――もあるが、ここでは理論と思想を包含するという意味での「経済学史」で統一する。なお、アメリカでは History of Economic Thought よりも History of Economics の表記（やはり理論史指向）が好んで使われるようである。

[9] さらに経済史まで考慮すれば、思想と実態（事実関係）の関連も重要となる。実態を反映する思想（例：唯物史観）、実態を動かす思想（例：プロテスタンティズムの倫理）、そし

何のためにこうした再構成を行うのかという未来に向けた知の営みも必要となり、すなわち「全体的再構成」economics for the future の視点が不可欠となる。つまりある未来志向の主観的な洞察力（シナリオ）を通じて、思想の歴史的展開を一望の下（パノラマ）に描き出す必要性である[10]。我々はこの３つの方法論――「合理的再構成」・「歴史的再構成」・「全体的再構成」――を十分に意識しながら、具体的な課題設定を通じて、経済思想の伝播を描いていく。

経済学史の研究者は、伝統的に地道なテキスト再構成[11]に携わってきた。現実の経済・市場社会を直接的に解釈・分析する経済理論家や、過去から現在の経済を――市場社会という時空に捕らわれることなく――直接的に解釈・分析する経済史家とは分業する形で、経済学史家・経済思想史家は《経済学者の眼》という間接的なメガネを通じて「経済」を眺め、経済学者の分析用具 rules of procedure を生み出す世界観 vision――「分析以前の認知活動」（Schumpeter 1994 [1951]: 41）――を発見・発掘することを生業としてきた。

経済学史家による解釈（再構成）が公表された時、通常科学と同じく、批判と競争とを通じてより良い認識が徐々に積み上がってきた場合も多いだろう。良質な研究者集団という装置そのものが、経済思想の望ましい再構成に寄与してきたと言えるのである。ただしこの再構成を実行するのは、いずれの場合でも現在の研究者である。再構成に当たって、我々自身の通念・世界観が反映され、時代の制約を受ける。どんなに広い視野や自省や批判的視座を持っていても、逃れられない盲点は存在する。

2-2　テキストマイニングとは何か

この盲点を顕在化させる１つの手段として、テキストマイニング text mining の採用を考えてみたい。定性的（質的）分析と分類される通常の経済学史研究を進めた後、いったんは個人の意志とは離れた量的分析をテキスト再構成の過程で噛み合わせるならば、全体としての解釈に合理性・一貫性・説得性を増加させることが可能になるのではないか。例えば、我々が重要とは見なさない「過

て実態と思想のズレ（理念の実現を妨げる過程は何か）という三態がありえる。特に３番目の緊張関係は重要である。小室（2012: 47-48）を参照せよ。

10 この３つの区分、および経済学史の「パノラマ=シナリオ・モデル」は塩野谷（2009: 350-351, 355）から借用した。ただし、英語部分は本稿独自である。

11 テキストは質的データそのものであるが、従来の質的分析はこの《テキスト》を「自由回答式 open-ended データ」と同義と考えてきた（Creswell 2015: 124／訳 141）。我々はここに「原典」という意味も加えている。

去の専門用語・日常用語」を通常のテキスト解釈では見逃しても、機械的に処理されたテキスト解析によって「新たなデータとして発見」することが可能になるかもしれない。ここにテキストマイニングという技法を採用し、従来の経済学史研究と接合させる可能性が見いだせる。

このテキストマイニング分析を、従来の統計分析 statistical analysis やデータマイニング data mining と比較対照することで説明しよう。

従来（1990 年代頃まで）の統計分析では、情報を得るための費用が高く、その情報を処理するための演算能力が低いという環境を前提としていた。その中で「できるだけ小さい情報量から、世界の姿を知ろうとする試み」（岡嶋 2006: 10）が統計分析であり、想定された法則の事後検証を得意としてきた。現実の経済行動に対して高い説明能力を有すると考えられている行動経済学あるいは実験経済学といった比較的新しい経済学においても、限られた情報量から結論を導出せざるをえない。

他方、データマイニングは次のように説明される[12]。

> 「データ内の情報や意思決定に使用される知識を特定するために用いられるさまざまな手法」（喜田 2008: 26）

データマイニングという概念は、新しい手法の開発というよりは、情報をめぐる環境の激変に促されて誕生した。すなわち、大量のデータ・情報が安価に入手可能で、その情報を処理する演算能力が圧倒的に高くなったという環境変化である。また、プログラミング言語やアルゴリズムの改善による強力なソフトウェアの開発と流布も圧倒的な力を持った（Ignatow & Mihalcea 2017: 5）。対象となる情報の質が格段に向上し、その量や範囲が飛躍的に増大した。その結果、大量の情報を一括して、短時間で処理できるようになり、分析の精度を格段と高めることが可能になった。古典的な統計学では小規模データの解析を目的として、例えばデータは正規分布に従ったり分散が等しいと見なしたりという複数の仮定が要求されていた。しかし、データマイニングでは日々大量のデータが更新されており、正規分布に従うという仮定はほぼ成立しない（石田・小林 2013: 2）[13]。

---

12 「数値データや名義データから属性の傾向や属性間の規則的な関係を統計的パターンとして抽出するアルゴリズムの総称」（村松・三浦 2009: 137）、という説明もある。

13 データマイニングが伝統的な統計学に対峙するというよりは、最先端で有力な一分野と

また、データマイニングの対象が数値や商品名という一意のデータに対し、テキストマイニングの対象は《テキスト》自体であり、文書内の記述は必ずしも一意ではない（那須川 2006: 19）。あるいは前者では表形式で集計された数値（構造化データ）を扱うのに対して、後者では加工されていない生の文章（非構造化データ）を扱う（小林 2017b: 11）。ゆえにテキストマイニングでは、データとしてのテキストを範疇ごとに分類したり、グラフや GUI によって可視化したりという別の作業が必要となる。処理の方法としては、回帰分析・決定木分析・クラスター分析・ニューラルネットワークなど多岐にわたる[14]。

マイニングとは元々「採鉱する」「坑道を掘る」という意味であり、データマイニングやテキストマイニングには２つの基本的な特性がある[15]。第１に、大量の情報から隠された法則や知見を抽出することである。第２に、このように抽出された法則の中から、意義のある、有意味な、有用な法則を厳選することである。第１の段階では主に統計処理ソフトやアルゴリズムの理解など、数理的な処理（量的分析）が助けとなる。しかし第２の段階では、自明ではないが有意味な法則・解釈・知見は何かという判断が不可欠であり、ここに質的分析も同時に求められていることに留意したい。すなわち、シミュレーションなどの予測を前提にした議論ではなく、あくまで分析対象における確実な傾向分析を展開することが要望されている。

さて、テキストマイニングはおおよそ 2000 年以降、急速に発展してきた分析手法である。ところが、この概念に一般的な定義を与えるのは難しい。最大公約数として「テキストからの知識の発見」（喜田 2008: 27）という説明もあるが、これだけでは従来の経済学史研究や思想史研究との差異を浮き彫りにすることはできない。定義が困難な理由の１つは、テキストマイニングに２つの方向性が混在しているため[16]である。

第１の方向性は、テキストマイニングをデータマイニングの一種、あるいは拡張と捉える理解である。データマイニングにおけるデータは典型的には数値であり（あるいは表形式として定型化されており）、この数量的なデータは「構

して位置づけられつつある（豊田編 2008: v）。

14 石田・金編（2012: 6-14）も主な方法を列挙している。中でも、「視覚化」を筆頭に挙げている。

15 岡嶋（2006: 28）の指摘による。

16 小林（2017a: 19）は２つの混在というよりも、テキスト分析は自然言語処理（形態素や構文の解析）とデータマイニング（頻度・統計処理・視覚化）の双方をこの順番で用いたものと捉えている。

造化されている」と呼ぶことができる。そこでテキストという自然言語を――文字・単語 などの単位に分解し――数量化・構造化するデータとして変換すれば、後は通常のデータマイニングの手法が適用できる。

第２の方向性は、テキストマイニングを自然言語処理 natural language processing と捉え、日常用語を分類や検索によって計算機が処理しやすい表現に直すことで、テキストデータ（特に、テキスト内部の最小単位である単語）から新しい知見を引き出す理解である。つまり、人工知能や言語学の一分野と捉えられる。この場合、テキストマイニングはデータマイニングと一線を画すことになる。

さらに、現在ではこの両者を織り交ぜた方法が数多く考案されており[17]、テキストマイニング独自の定義を示すことは難しい。実際、以下のようなやや漠然とした説明が加えられることも多い。

> 「小説、新聞、メール、日記、ブログ、報告書、演説文などは文字列によっ て自由に記述されたものであり、表形式に定型化されていない。このような文字列で記述されたテキストデータの山から情報や知識を探し出すことを目的とした分野」（金 2009: v）
>
> 「テキストデータを計算機で定量的に解析して有用な方法を抽出するための様々な手法の総称」（村松・三浦 2009: 1）[18]
>
> 「テキストの意味論的分析に向け、コンピュータに基礎付けられた方法であり、テキスト（特に非常に大量のテキスト）を自動的に（あるいは半自動的に）構造化する助けとなる」[19]

また、山本 (2011: 81) はテキストマイニングを、データ中心の分析 (統計解析) と人間中心の分析（フィールドワークや事例研究）の中間に位置する「文脈中心の分析」と捉えている[20]。

---

17 その一例として、喜田（2010: iii）は両者の利点を取り入れた「混合マイニ ング」を提唱している。

18 もう１つの説明として、「大量のテキストデータを統一的な視点から少ない労力で分析する（村松・三浦 2009: 1）、とある。

19 Heyer (2009)という文献を紹介した Wiedemann (2016: 2-3)の定義による。

20 末田ほか編（2011: 226）も同様に、元来、テキストマイニングが「量的研究と質的研究の中間的な位置」と見なしている。

以上を踏まえて、やや抽象的であることを許容するならば、テキストマイニングの意義としては、次の３つの説明が有用である。

> 「単なる検索や分類整理とは異なり、複数の文書データの内容を総合的にとらえることで初めて得られる知見を抽出するための内容分析の技術」（佐々木 2012: 226）
>
> 「テキストマイニングとは、検索および分類集計の範囲を越えた知識発見の実現を志向するもので、個々のテキストを読んだだけでは得られない知識を獲得する技術」（菰田・那須川編 2014: 25）
>
> 「テキストの内容の類似度をソフトウェアで判定してグループ化する…。さらには、ていねいに読んでも気づきにくい隠れた構造を探る…。」（石田基弘 2017: i）

いずれも大量のテキストデータを単に解析するのではなく、新たな知見を発見する[21]ための手段としてテキストマイニングを捉えている。この方法によってテキスト全体の性質および注目すべき特徴を数値と視覚によって捉えられ、テキスト解釈の過程――あるいは、解析の再現可能性――の明示化による信頼性を確保できる（末田ほか編 2011: 239; 田崎編 2015: 190-191）。

2-3　テキストマイニングの注目度

この手法への注目度を測るために、小木（2015）に倣って、論文検索サイト「CiNii」（http://ci.nii.ac.jp）において、フリーワード検索を試みた（2017.8.11 現在）。その結果は図 2-1 にあるが、小木（2015: 32）の推論とは若干異なった部分が出ている。まず、「テキストマイニング」を含む題名を持つ論文が初登場したのは 1999 年ではなく、1998 年である（その他、出版年不明の論文も 12 件ある）。次に、2011 年にピークを迎えたわけではなく、ジグザグは記録しながら、2015 年も 250 件弱になるなど、これからも漸増していく傾向が読み取れる。小木（2015: 32）は 2012 年の出版件数を 100 件としているが、現在の記録では 177 件となっている。2016 年の件数は 188 件だが、あと数年すればもっと多くの論文が登録されることになるだろう。

---

[21] 豊田編（2008: 18）は、発見された知識が持つ価値はそれを得るためにかけた時間と関係するとして、短時間で大量のデータを解析できるマイニングの効用を指摘している。

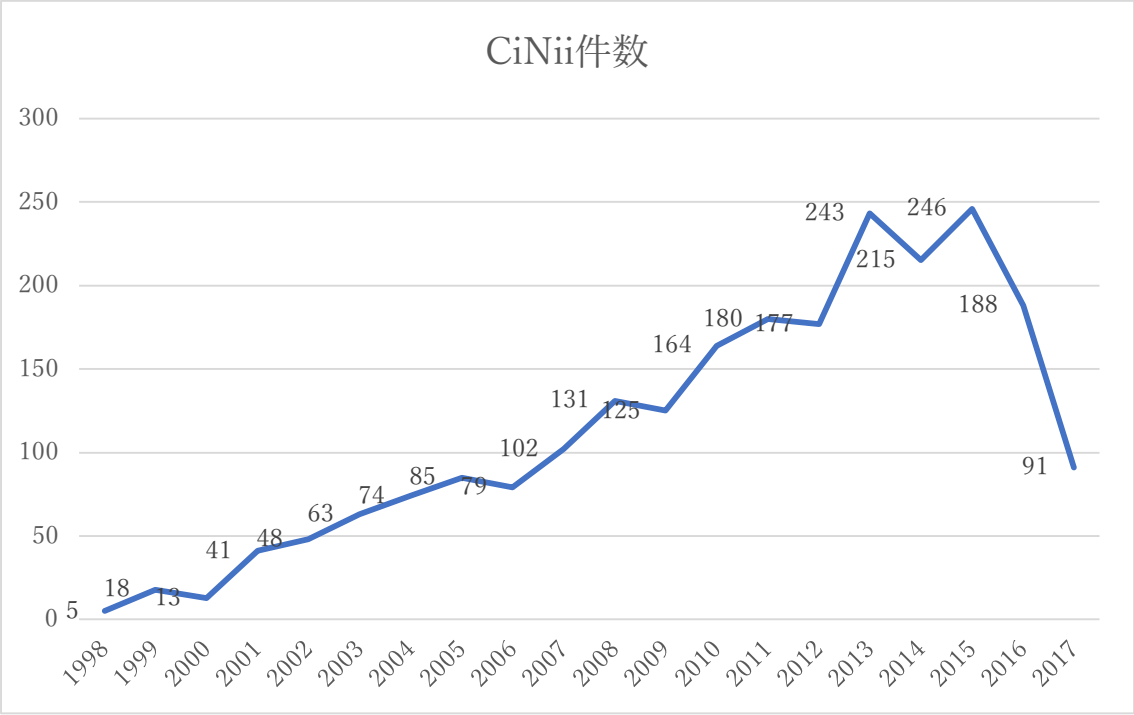


図 2-1　「テキストマイニング」を含む論文の出版件数

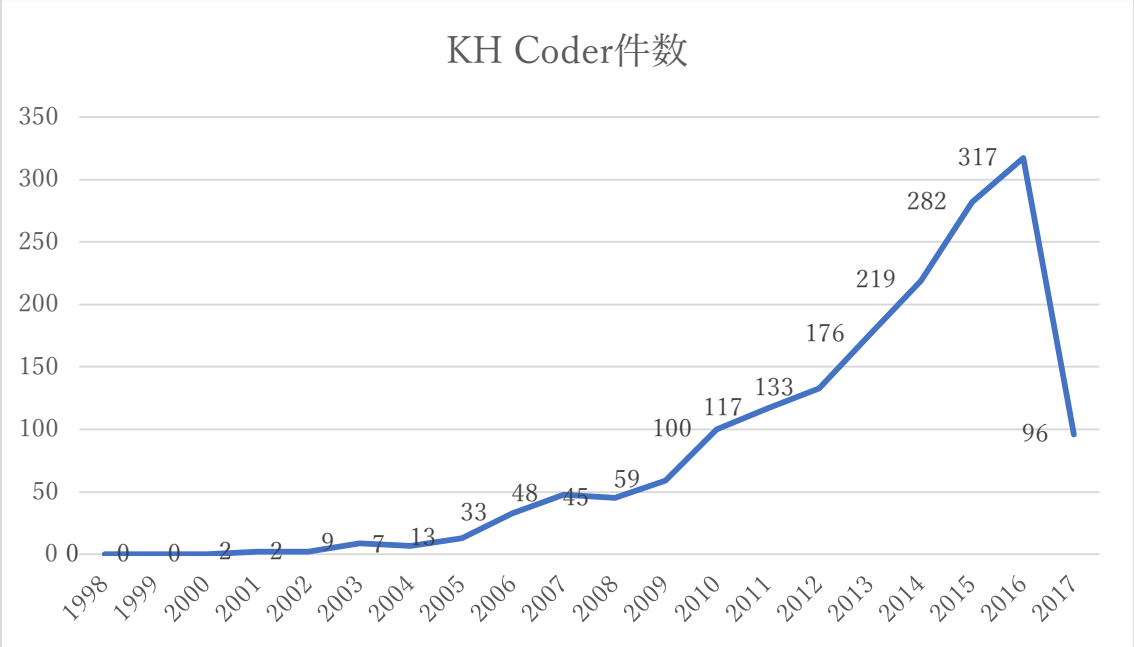


図 2-2　KH Coder を使用した論文の出版件数

全体的には、テキストマイニングを用いた論文の生産は、ますます伸びていると判断できる。小木（2015: 32）の指摘する通り、論文の公表と登録にはラグが存在する――この検索サイトは過去に遡って、漸次、論文を登録していく――ため、特に直近の増減傾向をつかむには細心の注意が必要となる。この例からわかるように、基礎となるデータの特徴を常に検証しなければならない。

第4節で導入するソフト（KH Coder）を使用した論文の出版件数も記録しておこう（2017.8.13 現在）。この書誌データベース[22]がどのように構築されているかは不明だが、論文の著者からソフト製作者への自発的な通知を含んでいると推測される。この書誌データベースには、「テキストマイニング」という単語を題名に含まないもののテキストマイニング分析を確実に用いている論文も多く収録されている。さらに、両者の図を比べてみると、2015 年以降は KH Coder 件数の方が常に多くなっている。2016 年も前年より増えている（図 2-2）。つまり、CiNii のデータベースから「テキストマイニング」のヒット数を記録するだけでは、テキストマイニング分析の広がりを適切に判断するには不十分となる。

Google Ngram Viewer による検索結果も紹介しておこう（図 2-3）。この検索エンジンはグーグル社がスキャンした 520 万冊の本（1500 年から 2008 年までの出版）にある情報（5000 億単語）[23]を N-gram（テキスト内にある隣接する言葉がどの程度、共起するかという指標）[24]を用いてグラフ化したものである。元になるデータは極めて限定されているが、「text mining」という言葉は 1994 年以降、ほぼ一定の割合で増加していたことがわかる。

---

[22] http://khc.sourceforge.net/bib.html?year=2017&auth=all&key=.　総数 1658 件。

[23] Mark Russell, “Google Database Tracks Popularity of 500B Words”, *Newser*, December 17 2010, http://www.newser.com/story/107766/google-database-tracks-popularity-of-500b-words.html (2017.8.12 アクセス)

[24] 「連続する文字ないし形態素、品詞をペアとした頻度情報」（石田・小林 2013: 68）。「文章における n 個の要素の連鎖」（小林 2017a: 126）。「雇用を確保する」という例文の場合、文字 2-gram ならば「雇 用」「用 を」「を 確」…、単語 2-gram ならば「雇用 を」「を 確保」「確保 する」という具合に、機械的に抽出される。

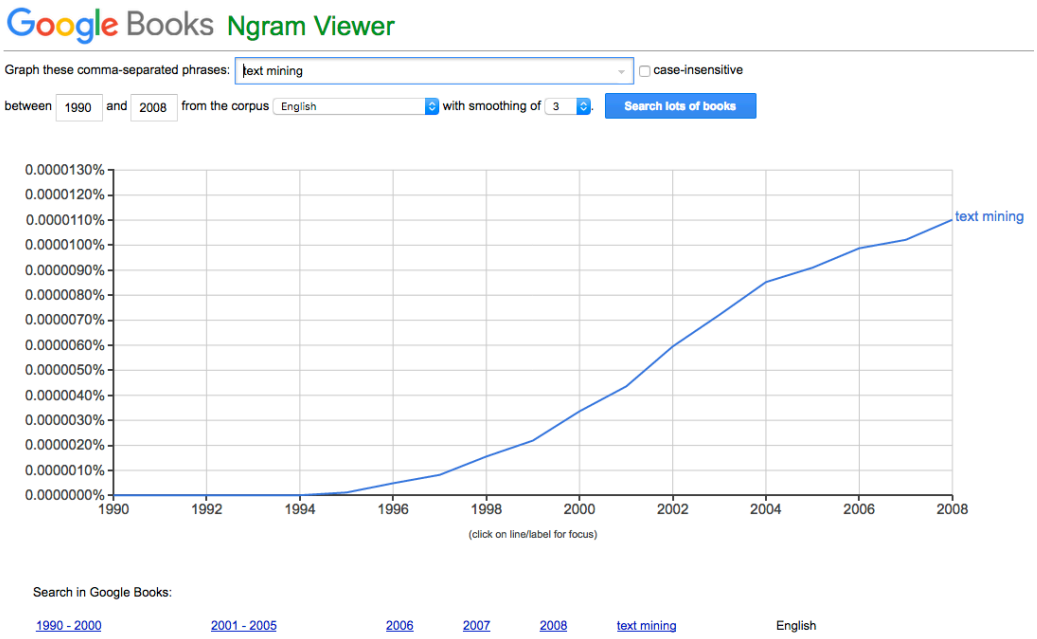


図 2-3　Google Ngram Viewer（‘text mining’で検索、2017.8.12）

現在、テキストマイニングは社会科学でも多くの分野で応用され始めている。アンケートや広告を読み解くもの、新聞記事や有価証券報告書を対象とするもの[25]、既存のデータベース（例：JST 失敗知識データベース）を用いたもの[26]、ネットの言説を分析したもの[27]など。この現象は、社会科学でも人文科学でもテキストこそが最も頻繁に研究されてきたデータであること（Wiedemnn 2016: 1）を考えれば、驚くべきことではない。ただし、社会科学への応用を推進する研究者も、ほとんどの場合、その例示を人類学・教育学・社会学・政治科学・経営学として（Ignatow & Mihalcea 2017: 3; Wiedmnn 2016: 1）、法学や経済学への応用を指向する研究はほとんど見当たらない。

経済学史研究のように、専門家が執筆したテキスト（一次文献）そのものを対象に取りあげているのは、管見の限りでは、我々の研究グループ以外では極めて珍しい。ただし、いくつかの先駆的な業績を次のように略述しておこう。まず Wright (2016)は社会的関係を視覚的に表現する「社会ネットワーク分析」

25 佐々木（2012, 2015, 2016）は言説分析のために新聞記事データベースを用いており、喜田（2007）は経営学の立場から有価証券報告書を分析した。
26 山本（2011）は経営学（知識工学）の立場である。
27 高（2015）は社会心理学の立場から、ネット上の現代的レイシズムを分析した。

[28]を用いて、1920 年代のウィーンを舞台に、多様な個人と集団が学問分野を超えて交流する様子を描いた。Claveau & Gingras (2016)は計量書誌学と動的ネットワーク分析を組み合わせて、1956 年から 2014 年までの期間で、経済学の内部でどのような専門分野が盛衰するかを描いた。Binder & Jennings (2016)は「トピックモデル」[29]という手法を用いて、スミス『国富論』に関して、キャナン版が付加した見出しに大きく影響された現代の解釈と、1784 年版の原典から機械的に抽出できる解釈を対比している。O'Neill (2016)は J. S.ミルが London Library から借り出した作品や寄付した書物などデータと、ミル自身の著作の網羅的語彙集 corpus とを対応させ、その直接的関係を導いた。以上のように、経済学史の文脈でも、テキストマイニングの適用が始まった段階である。

2-4　経済学史への適用

このようなテキストマイニングの手法は、従来からの経済学史研究にどのような意味を持つのだろうか。結論を先に言えば、二重の相互作用[30]によって、望ましい相乗効果が期待できるだろう、という意義を持つ。

第 1 の相互作用は、質的分析と量的分析の間に発生する（図 1-4 の赤矢印）。テキストマイニングの方法ではデジタル化されたデータ収集から始め、統計的手法を導入するための前処理を行い、そのデータを解析して結果を得る。特に、単語レベルの関係性[31]が瞬時に正確に抽出される。この過程で、各テキストは形態素（分かち書き）に分解され、文脈レベルでも単語レベルでも切り離され、いったん意味が喪失している（「二重の喪失」[32]）。つまり、テキストマイニングによる自動的な解析は、それ自体ではあまり意味を持たない。しかし、マイニングは技法・技術という機械処理だけで完結するわけではない。実際には、

---

[28] 社会科学で意義あるデータとして、①属性 attribute、②関係性 relational、③観念的 ideational という分類をしておけば（Scott 2013: 3）、通常の計量経済学は①に、社会ネットワーク分析は②に、我々の関心（意味論 semantics）は③にあると言えるだろう。

[29] 自然言語処理における潜在的意味解析 latent semantic analysis の文脈で発展している手法で、データの背後にある隠れたトピックを推定する方法である（小林 2017a: 179）。

[30] この型は Creswell（2015: 37, 39, 41／訳 42, 44, 47）が挙げる 3 つの型――収斂型デザイン・逐次説明型デザイン・逐次探求型デザイン――のいずれとも異なる。収斂型では当初から質的データと量的データを同時に収集して考察している。逐次説明型では量的データを説明するために質的データを導入している。逐次探求型では質的データの収集・分析の後に量的データが視野に入る。

[31] 佐藤（2008: 57）は、テキストマイニングの特徴（限界）として、基本的な分析ユニットが単語（テキストの最小単位）であることを指摘している。

[32] 菰田・那須川編（2014: 46）の表現から示唆された。

テキスト全体に対して、様々な段階で「解釈」を施さなければならない。すなわち、仮説、収集されたデータそのものの意義、カテゴリーの名前付け、結果の意義と限界、そしてテキスト全体の総合的解析という各場面で、解釈が重要な位置を占める。

換言すれば、いったん解体されたテキストを再構成する[33]ためには、解析者の観点・目的が必要不可欠なのである。その文書の意義をあらかじめ熟知している解析者は、テキストマイニングの前に「柔らかな作業仮説」を既に持っている。その観点から、次の（グループ分けや新しい文書との比較を含めた）分析方法が定まり、十分に精確なテキストマイニングが可能になり、「より固まった結論」を導ける可能性が高まる[34]。質的分析と量的分析は、言わばサンドウィッチのように挟み込まれた関係が必要となる。この意味で、解析者である研究者の内部に、質的分析と量的分析の相互作用・相互交流を意識的に発生させなければならない。

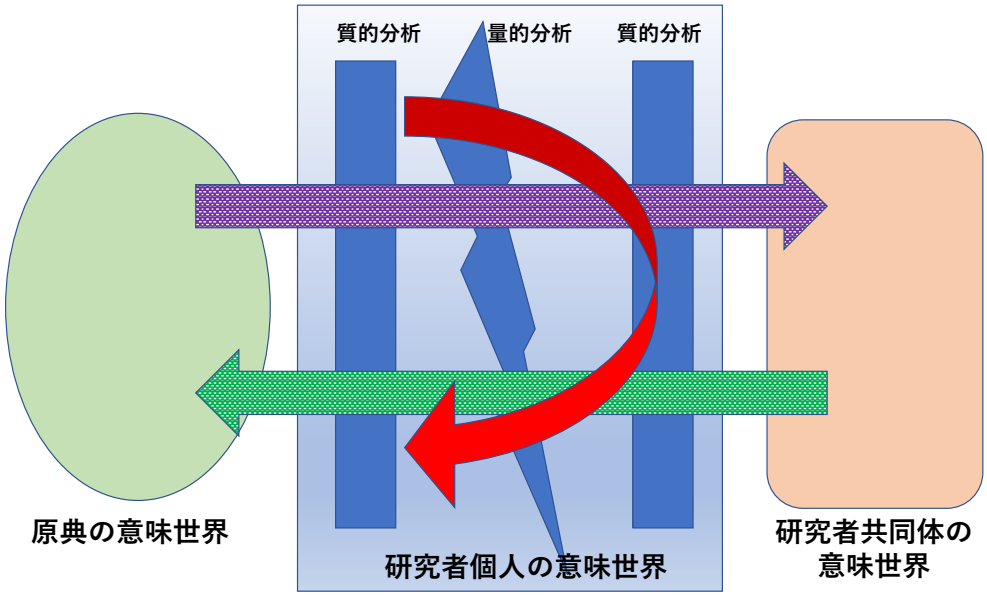


図 1-4　二重の相互作用

[33] 佐藤（2008: 50）は、この過程を文書セグメントの「再文脈化」と呼び、第１段階（データベース化；分類と配列）と第２段階（ストーリー化；新しい意味づけ）に分けた。

[34] 佐藤（2008: 12）は優れた質的研究のためには、手持ちのデータで全体の筋を粗いデッサンのようにいったん仕上げてみて、問題点を浮き彫りにさせた後、データの追加的収集と解析に進んでいくべき、と述べている。このような相互作用のため、解析技術を熟知した者とテキストの意義を熟知した者が、分業すれば効率的であろう。

第２の相互作用は、研究者個人を媒介として、原典と研究者共同体の間に発生する[35]（図 1-4 の紫右方向矢印と緑左方向矢印）。経済学史の研究は、過去の経済学者による《原典》が持っている個別の意味・意義・価値観・世界観・社会的重要性を、研究者個人の観点や分析を通じて、現在の学問という研究者の共同体に共有される言葉、つまり普遍的で抽象的な知・知見に変換していくことである――そして、研究者共同体の先に、我々自身が生きている社会に、その知見を届けることである。この変換は図の左から右への矢印の方向だけで行われるのではなく、必ず、研究者共同体のロジック・世界観からの方向性もある。２つの世界を研究者個人とその言葉とを媒介として、何度となく往復できる場合、「厚みのある記述」が可能になるのではないか。

こうしたテキストマイニングという新しい技術と従来の経済学の知見を交互に組み合わせ、相乗効果を発揮させれば、従来は見逃されていた新しい知見――例えば，匿名文書の筆者の特定、当該書籍の版別異同における思想的変化の明確化、通俗化・単純化された支配的言説の変形など――が生まれる可能性を秘めている。つまり、研究者の直観や長年の蓄積によって支えられてきた質的なテキスト解釈の途中に、大量のデータを統一的に処理するという量的解析を挟み込むことによって、個別的で特殊な事例報告を越え（石田 2008: 5）、より深みのある、説得力の増す仮説・知見を見出すことができるのではないか。さらに、この知見が原典と研究者共同体をさらに緊密に結びつけ、通常は顧みられることの少ない原典の現代における意義を再確認できるのではないだろうか。

この意味で、テキストマイニングは経済学史研究と適合的であり、より「発見的」heuristic な知見を促す新しい手法となりえよう[36]。そして、研究手法においても多様なアプローチがあることを示せれば、過去の経済学者たちの主張を、より客観的・効果的に掘り起こす手法が競合する余地を生じさせ、様々な分野の研究者が経済学史・経済思想史という分野に参入する契機となりうるだろう[37]。留意すべきは、この質的分析と量的分析は背反ではなく、むしろ相乗効

---

[35] ３つの意味世界という表現や図は佐藤（2008: 28）から借りたが、図の内容や意味づけは本稿と異なる。

[36] 村井編（2014: 140）も同じ立場。ただし、あくまでこの方法は期待される効能であり、逆に、テキストマイニングという手法が、従来の経済学史研究と比べて有効でないと判明する場合もありうる。この場合、旧来の計量文献学に代表されるように、匿名の文書・パンフレットの著者特定などという限定された領域において、この手法を経済学史研究に適用することになるだろう。

[37] 逆に、経済学史の研究者が他分野の手法を取得する良い機会ともなる。

果を持つ点である。テキストマイニングという新しい手法は、従来から蓄積してきた質的分析と結合することで、経済学史という分野自体の伝播を広げる可能性を持つだろう。

2-5（補）量的分析と質的分析

この項では本論からやや外れ、社会科学における量的（定量的）分析 quantitative analysis と質的（定性的）分析 qualitative analysis の差異や関係性[38]について、経済学史の方法と絡めながら略述しておこう。

定量的分析は、データセットの観察を任務とする。データセットとは従属変数（予測の対象となる量であり、目的変数）と独立変数（その変動を説明すると仮に考えられる量であり、説明変数）の両方を指す。大量のデータを一括して処理できるため、観察視野の広範さが確保される。数学的な基盤としては、推測統計（統計学と確率論）に依存している（Goertz & Mahoney 2012: 2／訳 2）。他方、独立変数の選び方に主観や仮定が入り込むため、いったん取りあげられた説明変数以外の要素を見落とす可能性もある。データ同士の質的な差は無視できる（等価値である）と仮定しており、個々の事例に関する知識は断片的となる。これは Brady & Collier eds. (2010: 180／訳 279)が呼ぶところの「薄い分析」[39]thin analysis である。ただし、質的な差を無視できるから大量のデータを一括処理できるのであり、「薄い分析」が劣った分析であることを意味しない。

対照的に、定性的分析の主流は因果過程の観察を得意とする[40]。事例研究 case study が典型であるように、ごく少数の事例だとしても、対象に対する洞察の深さによって、社会的事象の因果関係を推察することになる。他には、参与観察 ethnography[41]、インタビュー、会話（談話）分析 conversation analysis、言説

---

[38] Goertz & Mahoney (2012: 1／訳 1)は両者の方法を、異なった伝統と価値観を持つ２つの文化と捉えた。

[39] 佐藤（2008: 5）は、「厚い thick 記述」に対して「お手軽な quick 記述」、「密な記述」に対して「雑な記述」を対置している。その上で、７つの類型に分けられた「薄い記述」を避ける観点を提供している。村井編（2014: 152）はより中立的に、「深い」分析・「広い」分析と分けている。深さから広さに至るには、例えばデータの形式を整えることであり、広さから深さに至るには、例えばさらに細分化した定量データを包括的に収集することがありえる。

[40] 「質的研究は経験的資料として数ではなくテキストを用い、研究する現実が社会的構成物であるとの考えから始める」（Flick 2007: 3／訳 2）。

[41] 研究対象そのものに調査者が参加して、生活を共にしながら調査を続ける方法。民族誌（学）を意味するエスノグラフィーと類義語である。

分析 discourse analysis[42]などがある。これは事例や文脈に関する豊富な知識に基づき、洞察を深めることで、推論の妥当性を高める作業とも言える。定量的なモデルの中で考察される変数以外に、潜在的に重要な要因をあぶり出すことを得意とする。Brady & Collier eds. (2010: 355／訳 307)によれば「厚みのある分析」thick analysis であり、ある事象が個々人にとってどのような意味を持つかという解釈学的・哲学的な問いともなる。数学的な基盤としては、論理学と集合論に立脚している（Goertz & Mahoney 2012: 16／訳 19）[43]。主観確率であるベイズ主義[44]を採用し、情報不足に起因する不確かさを、主観的な信念の度合いとして定量化する試みとなる。

定性的分析の傍流は、テキストの内的論理・外的文脈という「意味」を重視する解釈的方法にある。この潮流は一般化が難しいために、定性的分析を重視する論者でも「定性的研究パラダイムにもさまざまな内部分裂が存在する」（Goertz & Mahoney 2012: 4／訳 5）として、この部分を無視することが多い。しかし、経済学史研究はこの方法に最も合致している。1-1 で「合理的再構成」「歴史的再構成」「全体的再構成」という３つの過程を取りあげたように、この分野の分析方法について、より深い洞察が求められる。

多くの論者が述べているように、量的分析と質的分析はそれぞれ価値観や手続きが異なるものの、「多少プラグマティックな折衷主義」（Flick 2007: 7／訳 9）が用いられ、「緩やかな垣根を越えて互いに影響を与えることもできる」（Goertz & Mahoney 2012: 226／訳 257）し、「両者の関係は循環的なもの」（樋口 2014: 9、別の文献からの引用）、「相補的な関係にあり、それだけでなく相乗的な関係にある」（佐々木 2016: 246）。両者の接合[45]という場面で、経済学史への適用が視野に入ってくる。

第３節　テキストマイニングの手順

---

42 会話分析と言説分析の差異は、前者が会話という個人的関係に注目するのに対して、後者がある言説とその社会的文脈との影響関係を探る点にある。野村（2017: 252, 256）を参照。なお類似の概念に内容分析 content analysis もある。これは「文章・音声・映像などさまざまな質的データを分析するための方法」（樋口 2014: 1）であり、記述のみならず推論も含む。

43 実際、石田淳（2017: 77）はブール代数（論理・集合演算の抽象モデル）を用いて、少数の事例から因果関係を把握する「質的比較分析」の洗練化を行っている。

44 対立する概念が頻度主義であり、不確かさをランダム性によって処理する。

45 「定量的あるいは定性的アプローチのどちらか一方のみに固執すれば、科学の進歩を妨げることになる」（Brady & Collier eds. 2010: 109／訳 203）。

本節の 3-1 と 3-2[46]では下平・小峯・松山（2012）で行った分析過程をやや詳しく記述する。3-3 では単語の「類似度」について一例を挙げる。経済学史研究ではテキストマイニングの手法に馴染みがないため、その分析過程においても詳細な検証が必要だからである。また、本稿ではその時に用いたソフトや分析方法と異なる部分が多く、双方を比較するために再録の形をとる。

3-1　前処理の重要性

「前処理」はその名称とは異なって、分析の周辺にあるものと捉えるべきなのではなく、形態素解析や構文解析を実行しているので、計量的なテキスト分析に不可欠の部分である（田崎編 2015: 199）。経済学史研究にテキストマイニングを導入する際、次の４つの手続きが最低限は必要である。

(i)対象テキストを電子化し、OCR（optical character reader）処理をする。
(ii)OCR 処理されたテキストを Microsoft Word などのソフト上で整形[47]（文字化けを直し、無意味な記号を排除するなどの手作業を行うこと）し、テキストファイルとして保存する。
(iii)保存されたテキストファイルを Microsoft Excel に読み込ませ、タグ付けを行った後、CSV ファイルとして保存する（この項目は削除可能）。
(iv)テキストマイニングのツールにて解析する（図 3-1）。

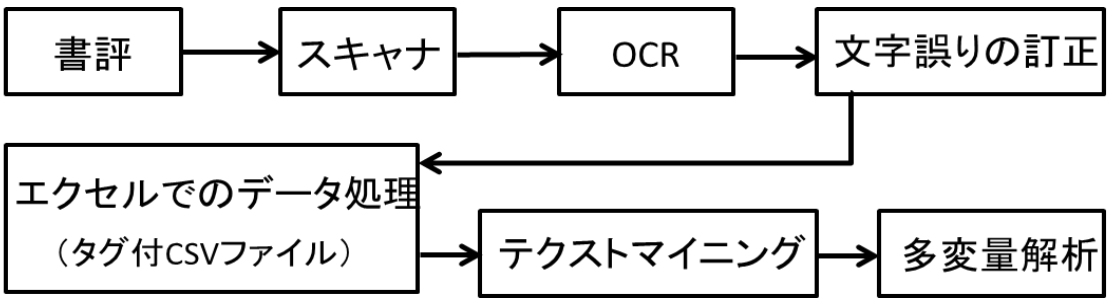


図 3-1　テキストマイニングの流れ

[46] 本節は、下平・小峯・松山（2012）の 2-1, 2-2, 3-1, 3-2, 3-3 を改訂したものである。
[47] 例えば、不要な改行を削除したい場合、「検索」（置換前）に「¥n（Mac OS の場合は¥r）」を入れ、「置換」に何も入れない、という操作をする。テキスト整形について詳しくは、小林（2017a: 41-46）を見よ。

テキストマイニング分析を行う場合、テキストそのものを解析処理しやすいように、前処理を行う必要がある。まず当該のテキストをデジタル化しなければならない。本研究では、Adobe 社製の Acrobat Professional というソフトを用いて、デジタル化されたテキスト（PDF ファイル）に対して、OCR（光学式文字読み取り装置）を施した。この処理によって、デジタル化されたテキスト全体を検索可能な文字列として認識させることができる。

次に、OCR にかけた当該テキストを Word ファイルに変換し、検索・置換の操作をする。この操作は、テキスト全体を一文一文のまとまりに切り込むためである。まず、Word のメニューバーより「検索」ウィンドウを表示させる。ウィンドウ内の「特殊文字」タブにて「段落記号」を選択することによって、置換後の文字列が「^p」と表示される。そして、検索する文字列をピリオド「.」、置換後の文字列を「.^p」とすれば、当該テキスト全体をピリオドの直後に改行することができる。この操作によって、テキスト全体がすべて一文ずつ改行された状態になる。

さらに、この文書をテキストエディタなどで読み込み（例：mi, 2.1.8 というテキスト・ソフトを使用）、Excel を開き、タブ区切りとして読み込ませる。この段階で、セルの一行一行に元文章が一行ずつ対応して表示される（図 3-2）。この各行「タグ付け」を行えば、その文書を管理することが容易になる（同じ書評で、節やページごとなどの別のタグを付けることも可能である）。最後に、この Excel ファイルを「カンマ区切り形式（CSV）」として保存する。汎用性のあるテキストマイニングのソフトの多くが、CSV ファイルに対応しているためである（第 4 節で用いるソフトはテキストファイルにも対応している）。

このようにして、やや煩雑ではあるが、テキストマイニング分析をするための下準備を行う。この部分を疎かにすると、分析の結果が大きく異なってくるので、細心の注意を払うことになる。特に OCR の精度を上げることは重要である。OCR の精度が低いと、整形に多大な時間がかかり、また検索が不正確になり、分析結果の根幹が揺らぐことになる。

下平・小峯・松山（2012）では、テキストマイニング分析を行うにあたって、TTM（TinyTextMiner v0.80）というフリーウェアを用いた[48]。これは CSV 形式の「タグ付きテキスト」を読み込んで、6 種類の集計データを自動的に作成するソフトである（図 3-2）。その 6 種類とは、語のタグ別集計（出現頻度）、

[48] http://mtmr.jp/ttm/, 作成者は村松真宏・三浦麻子の両名である。

単語のタグ別集計（出現件数）、語×タグのクロス集計（出現頻度）、語×タグのクロス集計（出現件数）[49]、語×語のクロス集計（出現件数）、テキスト×語のクロス集計（出現頻度）である。ここで「語」とは形態素 morpheme と呼ばれ、意味を持つ最小の文字列のことである。文を単語ごとに分割し、品詞情報などを付け加える作業を形態素解析と呼ぶ[50]（金 2009: 26）。日本語で、文を単語に分割することを「分かち書き」と呼ぶ（石田 2008: 8）。このソフトでは、動詞・名詞・形容詞・副詞を区別することが可能で、また日本語と英語が解析できる。

下平・小峯・松山（2012）では、元の文章を整形する際に、書評の本文のみを扱い、脚注は削除した。またページや題名など各頁におけるヘッダー、フッターの情報もすべて削除した。この作業に実は多くの労力が割かれる。

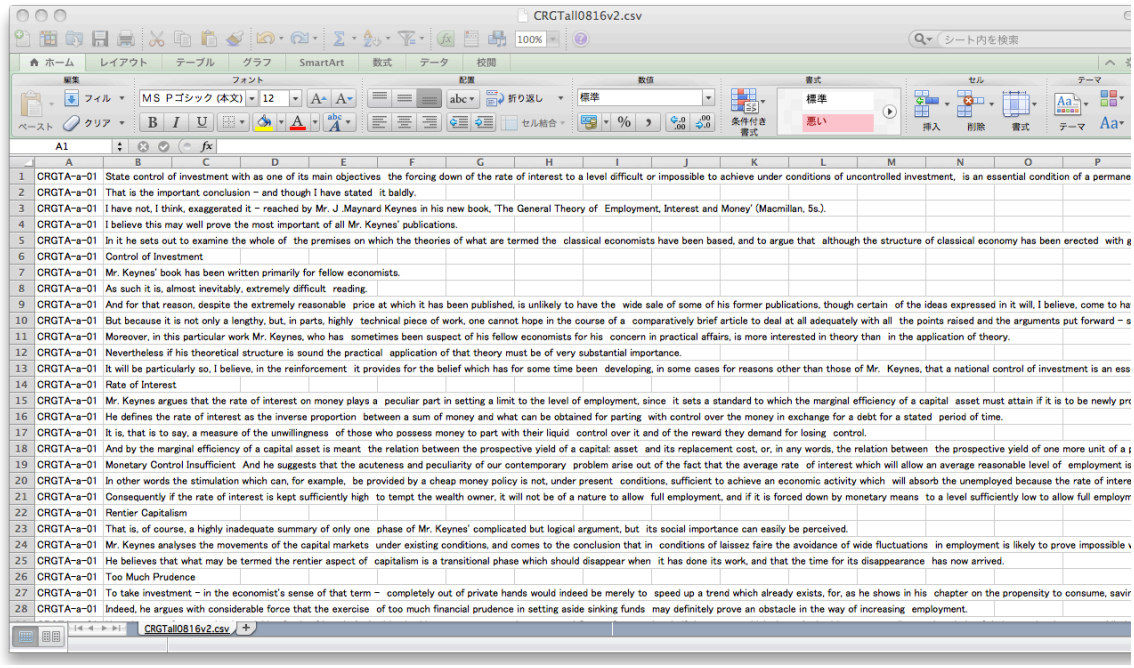


図 3-2　タブ区切りした Excel ファイル

3-2　頻度分析と多変量解析

前処理を終えた段階で、このソフトで出現頻度の高い名詞を抽出した。認知心理学でのカテゴリー化研究において、ある理念・概念・思想がもっとも象徴

49 同じ一文のテキストの中で、ある語が何回も出現した場合、出現頻度（のべ回数）ではすべてを数えるが、出現件数としては１つとなる。

50 分析対象のテキストが日本語の場合、分かち書きがなされていないので、形態素に切り出す作業が重要となる。村松・三浦（2009: 126）も参照。

的に現れるのが名詞に他ならないからである（喜田 2008: 151）。ただし、意味を持つと思われない高頻度単語（例：ケインズ『一般理論』における書評の場合、Mr, Keynes, book, economists など）はあらかじめ除外した（この作業において、有意味さの判定という価値判断を行っていることに留意）。

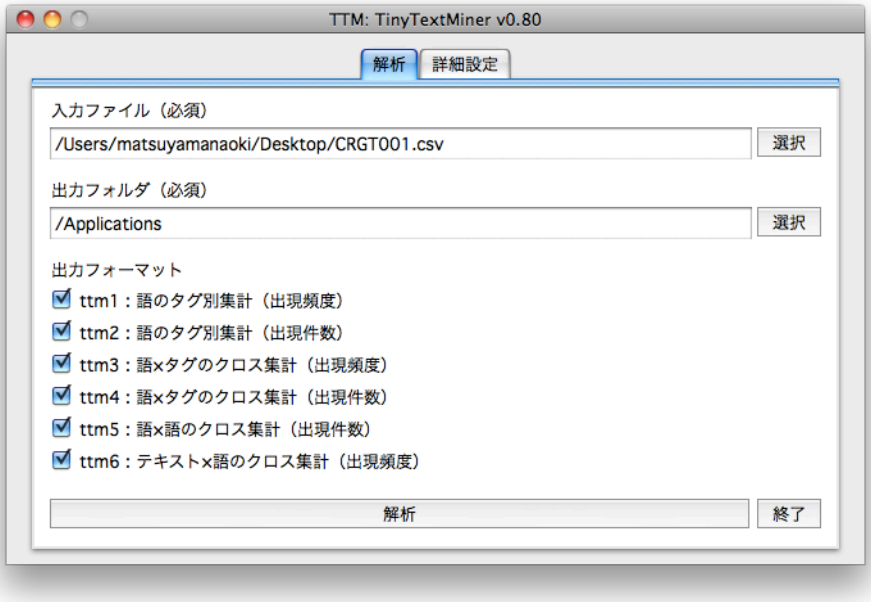
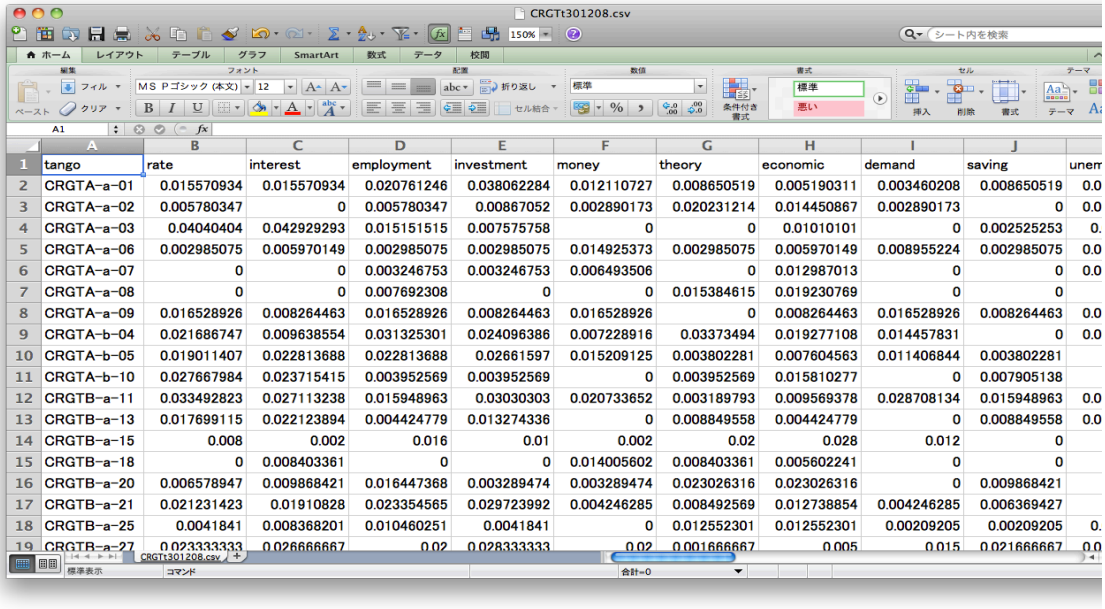


図 3-3　Tiny Text Miner に Excel ファイルを読み込ませる

出現頻度分析だけでは、発表媒体や執筆者に関する範疇別の差異がまだはっきりしない。そこで、次の段階として多変量解析という別の統計手法を用いることにする。多変量解析 multivariate analysis とは、複数の変数間に存在する影響・相互関連を一括して解析し、隠された傾向を探る統計的手法の総称である。主成分分析・対応分析・因子分析など、数多くの方法がある。いずれの場合も、複数の変数のデータを、何らかの方法で情報の損失を抑えながら少ない変数に集約したり分解したりすることになる。この抽出された変数こそ、多数のデータに埋もれた知見となりうる。その際、各データの平均値だけでなく、むしろ偏差（平均値からの離れ具合）を重視することになる。この偏差は各データの「個性」「ばらつき具合」を示すからである（涌井・涌井 2011: 19）。統計量としては分散（偏差の二乗平均）や共分散に注目することになる。

本節では統計ソフトとして R を用いた。R は統計計算とグラフィックスのための言語・環境であり、次のような特長を持つ[51]。１．無料であること。２．複数の OS に対応していること（Windows, Mac, Linux, Unix）。３．きめ細かなグラフィックスを作成できること。４．特定の機能に限定せず、使用者が新しい関数を拡張できること。５．既に複数パッケージが用意され（例えば R Commander は R の機能を GUI 方式で利用可能）、利便性が高い。

以下では、R Commander を用いた主成分分析を行う手順を紹介する。



CRGTt301208.csv

	A	B	C	D	E	F	G	H	I	J	
1	tango	rate	interest	employment	investment	money	theory	economic	demand	saving	unem
2	CRGTA-a-01	0.015570934	0.015570934	0.020761246	0.038062284	0.012110727	0.008650519	0.005190311	0.003460208	0.008650519	0.0
3	CRGTA-a-02	0.005780347	0	0.005780347	0.00867052	0.002890173	0.020231214	0.014450867	0.002890173	0	0.0
4	CRGTA-a-03	0.04040404	0.042929293	0.015151515	0.007575758	0	0	0.01010101	0	0.002525253	0.
5	CRGTA-a-06	0.002985075	0.005970149	0.002985075	0.002985075	0.014925373	0.002985075	0.005970149	0.008955224	0.002985075	0.0
6	CRGTA-a-07	0	0	0.003246753	0.003246753	0.006493506	0	0.012987013	0	0	0.0
7	CRGTA-a-08	0	0	0.007692308	0	0	0.015384615	0.019230769	0	0	
8	CRGTA-a-09	0.016528926	0.008264463	0.016528926	0.008264463	0.016528926	0	0.008264463	0.016528926	0.008264463	0.0
9	CRGTA-b-04	0.021686747	0.009638554	0.031325301	0.024096386	0.007228916	0.03373494	0.019277108	0.014457831	0	0.0
10	CRGTA-b-05	0.019011407	0.022813688	0.022813688	0.02661597	0.015209125	0.003802281	0.007604563	0.011406844	0.003802281	
11	CRGTA-b-10	0.027667984	0.023715415	0.003952569	0.003952569	0	0.003952569	0.015810277	0	0.007905138	
12	CRGTB-a-11	0.033492823	0.027113238	0.015948963	0.03030303	0.020733652	0.003189793	0.009569378	0.028708134	0.015948963	0.0
13	CRGTB-a-13	0.017699115	0.022123894	0.004424779	0.013274336	0	0.008849558	0.004424779	0	0.008849558	0.0
14	CRGTB-a-15	0.008	0.002	0.016	0.01	0.002	0.02	0.028	0.012	0	
15	CRGTB-a-18	0	0.008403361	0	0	0.014005602	0.008403361	0.005602241	0	0	
16	CRGTB-a-20	0.006578947	0.009868421	0.016447368	0.003289474	0.003289474	0.023026316	0.023026316	0	0.009868421	
17	CRGTB-a-21	0.021231423	0.01910828	0.023354565	0.029723992	0.004246285	0.008492569	0.012738854	0.004246285	0.006369427	
18	CRGTB-a-25	0.0041841	0.008368201	0.010460251	0.0041841	0	0.012552301	0.012552301	0.00209205	0.00209205	0.
19	CRGTB-a-27	0.023333333	0.026666667	0.02	0.028333333	0.02	0.001666667	0.005	0.015	0.021666667	0.0

図 3-4　R Commander 読込用に整形した CSV ファイル

まず、一列目にタグ、一行目に変数（この場合は単語）を配置した CSV ファイルを準備する。次に、R を立ち上げ、コマンドとして library(Rcmdr)を入力し、パッケージから R Commander を呼び出す。起動したら、上記の CSV ファイルをコピーする。「データのインポート」「テキストファイルまたはクリップボード、URL から」を選択することになる。「データファイルの場所」はクリップボードとして、「フィールドの区切り番号」は空白としておく。ここまでの作業は「データセットを表示」で確認できる。プルタブで、「統計量」→「次元解析」→「主成分分析」を選択する。「変数」は全てを選択し、「データセットに主成分得点を保存」を選ぶ。保存する主成分数は全て（この場合は

[51] http://www.okada.jp.org/RWiki/?RjpWiki には、インストールから掲示板まで、日本語環境による R に関する情報が集積している。

30 個）を選択する。ここまでの作業は、「出力ウィンドウ」で主成分・寄与率・累積寄与率等が表示される。また、「データセットを表示」で、元データの後ろに主成分得点が追加される。最後に、「出力をファイルに保存」で、出力ウィンドウがテキストファイル形式で保存される。「データ」「アクティブデータセット」「アクティブデータセットのエクスポート」を選択し、「行名をつける」チェックを外す。保存先を選択すると、テキストファイルとして出力される。その後、エクセルにデータをインポートすると加工が容易になる。

多変量解析の中でもクラスター分析 cluster analysis とは、測定された対象間の近さ（関連性）を基準に、対象をいくつかのグループ（クラスター）にまとめる統計的手法[52]である（村松・三浦 2009: 64）。あらかじめ分類や範疇を定めて行うグループ化と異なり、人間の予断なしに、時間もかかるラベル付与なしに、観測データから機械的に分ける方法である（奥野ほか 2016: 198）。「教師なし学習」unsupervised learning（指導を受けていない学習）と呼ばれることもある（小林 2017a: 169）。ここではその中でも、ある基準で対象間の類似度を測定し、似たもの同士を段階的にまとめていく階層的クラスタリングを採用した。この手順は主に次の 4 つである。

(i)データ集合の中から、互いの距離（類似度）[53]がもっとも近くなる（大きくなる）データのペアを捜す。
(ii)そのペアを 1 つのクラスターとして統合する。
(iii)そのクラスターと残りのデータ集合の中から、再び互いの距離がもっとも近くなるデータを探し、さらに 1 つのクラスターとして統合する。
(iv)データ全体が 1 つのクラスターに統合されるまで、上記の手順を繰り返す[54]。

クラスター分析では n 次元ベクトルの距離を測るのだが、現状では、様々な「距離」そのものに関する様々な考察が展開されており（金 2009: 161）、分

[52] 別の定義では、与えられた外側の基準を用いずに、対象の分散・相関・類似度・距離の情報を用いてグループ分けする分析である（金 2009: 160）。
[53] 距離や類似度は別に定義されている。ピアソン相関係数、コサイン類似度、ユークリッド距離など、複数の指標がある（金 2009: 161）。
[54] この手順の表現は http://gihyo.jp/dev/feature/01/visualization/0002 を参考にした。

析に応じてやや便宜的に変更されている状態である。このような距離の意味づけを経済学史的に確定する解釈を今後も続けていく必要があるだろう。

そこで、別の多変量解析である主成分分析[55]による解析を試みよう。主成分分析[56]principal component analysis とは、「多くの変数によって記述された量的データについて、複数の変数間の相関（共分散）を少数個の合成変数（これを主成分という）に縮約し、データの解釈を容易にするための分析手法」（村松・三浦 2009: 60）である。元の変数が n 個あれば、n 次元の情報があると言えるが、その情報の大半をごく少数の主成分に要約して表現できれば、この解析は成功したとみなせる。n 次元をすべて考慮するよりも、２つ３つの主成分のみを熟慮することができるし、また散布図 scatter plot[57]などによって、全体の傾向を視覚的に表現することも可能となる。少数の合成変数を作るには、多くの変数にウェイト（重み）という係数をかければよい。この場合、ウェイトの付け方は、合成変数が元の変数の情報量を、できるだけ多く含むように工夫しなければならない[58]。その基準としては、「１つ１つの個体が最もバラバラになるような変量の和を作る」（涌井・涌井 2011: 74）、つまり分散が最大になるような変量の和を求めれば良い。主成分係数とは主成分を作る際の係数を表し、その主成分がどのような特性を持っているかを示すので、ここに主観的な解釈を付与する余地がある。

主成分分析で得られる指標は主に４つある。第１に、固有値 eigen value である。固有値とはベクトルを一次変換した場合、元のベクトルの何倍になったかを示す値である。この分析における固有値は主成分の分散に対応しており、その主成分がどの程度、元のデータの情報を保持しているかを表す。第２に、寄与率 contribution である。ある主成分の固有値が表す情報が、すべてのデータの中で、何％ぐらい説明できるかという割合の指標[59]である。第３に、累積寄与

---

55 Amable (2003: 80／訳 143)は資本主義の多様性という理論を実証するために、主成分分析とクラスター分析を用いた各国の類型化を試みた。

56 別の説明では、「データの分散共分散、あるいは相関の情報にもとづいて多くの変数のデータの情報の損失を抑えながら少ない変数に集約して分析する方法」（金 2009: 147-148）とある。

57 ２種類の項目を縦軸と横軸にとり、１つの要素を打点 plot として記入する図のこと。相関関係を見いだすことができる。

58 データが２変数で２次元の散布図で示される場合、データの分散がもっとも大きくなる方向に軸をとると、これが第一主成分となる。データの散らばり具合が情報量だからである。次に平均値（重心）を通って、第１主成分である直線に直交する（無相関とする）線が第２主成分となる。

59 「資料全体の分散に占める主成分の分散の割合」（涌井・涌井 2011: 81）。

率とは、各主成分の寄与率を大きい順に足して累積した率である。そこまでの複数の主成分で、元のデータが持っていた情報量が全体として、どのくらい説明されているかを示す。第 4 に、主成分得点 principal component score である。これは主成分に個々のデータを代入したものであり、各データ（現在の分析では各書評）の特徴を検討する際などに用いられる。

ここで何番目までの主成分を用いるべきかが問題となる。一概には言えないが、目安として累積寄与率が 70〜80%、つまり全体の情報の 7 〜 8 割が確保されていれば良いという考え方がある（金 2009: 150; 村松・三浦 2009: 134）。あるいは主成分の固有値が 1 を超えれば良い。また各主成分の大きさを折れ線グラフに記入し（スクリープロット）、折れ線の傾きがゆるやかになる前までの主成分を採用する場合も見られる。

3-3　単語の類似度

4-3 では Jaccard 係数を用いた距離（類似度）を考察するが、ここでは岡崎（2016）に依拠して、別の類似度によって単語の近さを例示してみよう。基本的発想は、分布仮説 Distributional Hypothesis、つまり似た文脈で出現する単語同士は意味も似ているという考えである（奥野ほか 2016: 52）。

	have	new	drink	bottle	ride	speed	read
beer	36	14	72	57	3	0	1
wine	108	14	92	86	0	1	2
train	291	94	3	0	72	43	2

表 3-1　単語文脈行列

表 3-1 はある文書において、各単語の周辺（前後のいくつか）に出現する単語を例示したものである。例えば、wine という単語の周辺には bottle という単語が 86 回出現（共起）している。この場合、

beer (36, 14, 72, 57, 3, 0, 1)
wine (108, 14, 92, 86, 0, 1, 2)

という具合に、周辺に共起する単語の頻度を要素として、各単語がベクトル表示されていることになる。表 3-1 のような３行７列の行列を単語文脈行列と呼ぶ。単語をベクトル表示することで、その類似度を数学的に定義できることになる。ここでベクトルの余弦 cosine を考えると、２つのベクトルの向きが近いほど、そのなす角度 θ は小さくなる（cos θ は１に近づく）。そこで次のようなコサイン類似度を定義することができる。

$$\cos(\text{beer}, \text{wine}) = \Sigma\ (\text{beer}_i \cdot \text{wine}_i) \ /\ (\Sigma\,\text{beer}^2{}_i \cdot \Sigma\,\text{wine}^2{}_i)$$

上記の表に従ってこの値を計算すると、beer と wine の類似度は 0.941、beer と train は 0.387 となる（岡崎 2016: 52）。３つの単語の意味上の近さ・遠さが数値表現されたことになる。このように、テキストの処理にはコサイン類似度もよく用いられている（金 2009: 161；石田・金編 2012: 12）。

以上で、単語の類似度と下平・小峯・松山（2012）における解析方法の解説は終えた。次の節では、別のソフトと方法による分析を見ていこう。

第４節　ベヴァリッジ『自由社会における完全雇用』のテキスト解析

この節では、『一般理論』の政策への影響を測る１つの手段として、同時代の非常に有力な思想家による解釈を挟み込む。具体的には、ベヴァリッジの『自由社会における完全雇用』（1944）にケインズ的な要素がどの程度、存在するかをテキストマイニングの手法を借りて分析する。以下、最初の項では時代背景を説明し、次に実際の分析過程を頻度分析・階層的クラスター分析・共起ネットワーク分析・特徴語分析・複合語分析の順番で示す。

4-1　1930-40 年代の時代背景

ケインズが『一般理論』を世に問おうとした時、「すぐにではないのだが、次の 10 年間の間に、世界が経済問題を考える様式をおおよそ革命的に変えてしまうような経済理論に関する本を書いている」[60]と意気込んだ。ただし、実際には「本書は主として、私の仲間である経済学者たちに向けて書かれたものである」（Keynes 1936: v）と宣言されたように、直接の標的は経済学者という理

[60] Keynes (1973, vol. 13: 492), a letter to G. Barnard Shaw, 1 January 1935.

論家にあった。ある思考・理念が普及する過程の時間を考えると、理論（専門家集団に流布）と思想（一般大衆に流布）の中間に、《政策》（政策関与者に流布）が位置すると考えられる。そこで、政策に対する『一般理論』の浸透度、つまり「政策におけるケインズ革命」をどのように捉えたら良いだろうか。

より実践的な『貨幣改革論』（1923）や『貨幣論』（1930）と異なり、『一般理論』（1936）は内容的に直接的な政策指向ではないから、出版して数年経っても、政策担当者にはその革命的な思考はなかなか到達していなかった[61]。変化の契機は、イギリス経済の現況――開戦を控えて完全雇用に近づいてきたこと――と、戦時経済への速やかな対処の必要性であった。ケインズは 1939 年 11 月に「戦費調達論」の最初期版を公開し、戦費調達・インフレ回避・生活水準確保という三重の目的を「繰り延べ払い」という新機軸で達成しようと各方面を説得し始めた。この計画には国民所得の綿密な推計が不可欠であり、政府内外の声も受けて、イギリス政府はこの声に一定の反応をせざるを得なくなった。それが 1941 年の白書[62]および予算である。ここでインフレ・ギャップという概念――超過需要にある現実の国民所得と、潜在的な均衡国民所得との差――が用いられ、複式簿記の原則から各部門――政府や民間の所得・支出、産出、貯蓄・投資など――の数字が合一されることになった。このような白書と予算を、ケインズは「公共財政における革命である」[63]（Keynes 1979, vol. 22: 354）と激賞した。

政府高官や経済助言官の中には、経済や雇用の制御問題をもっと広い視野で考える者もいた。例えばケインズは労働大臣ベヴィン（労働党）に賛意を示し、「社会保障が戦後の国際政策の第一目標であるべき」[64]という戦争目的のメモを残していた。さらにこの時期のケインズは、次の認識を持っていた。

> 「戦時予算の重要性は、その予算が戦費を調達するからではない。…その重要性は社会的である。つまり、現在および今後、インフレという社会的な害悪を回避するため。社会的正義という一般的な良識を満たす形でそう

61 「ケインズ経済学への大蔵省の転換は、大戦が勃発した時にはまだ不完全だった」（Peden 1979: 61）。

62 *An Analysis of the Sources of War Finance and an Estimate of the National Income and Expenditure in 1938 and 1940*, HMSO, Cmd 6261, April 1941.

63 Keynes (1979, vol. 22: 354), a letter to F. A. Keynes, 14 April 1941.

64 TNA, PREM 4/100/5, “Professor Keynes’ Memorandum on War Aims”, 13 January 1941.このメモは閣内で閲覧された。

するため。他方、労働や経済に適切な誘因を持たせておくため。」（強調は原典）[65]

この場合、ケインズにとって「社会正義」とは、労働者の最低限所得保障・家族手当などを含む社会保障である。ゆえに、1942 年３月にベヴァリッジが戦後の社会保障政策に関連する広範な経済問題を含めてケインズに助言を求めた時、ケインズは「極めて重大で雄大な建設的計画である」[66]として、自らの方向に誘導しつつ、その助言に最大限に応えようとした。

保険と失業の専門家であるベヴァリッジは、政府から請われて委員会を 1941 年６月に発足させていた。しかし、政府の白書にもかかわらず、各省庁から派遣されていた中堅官僚は――政府側の思惑（労働災害補償などの限定された改善案）を超えて、大規模な予算を要求する包括的計画に改変されたため――1942 年１月までに全員が引きあげた。その結果、議長のベヴァリッジのみが署名する形で、1942 年 12 月に『社会保険および関連サービス』が公表された。前述のように、内閣経済部（主にミード）やケインズの助力も得つつ、戦後社会の青写真を描いた『ベヴァリッジ報告』がここに完成したのである。

一般大衆に熱狂的に受け入れられた青写真ではあったが、チャーチル内閣と政府高官の多くには、かなりの部分、不評であった。軍事的な戦争に注力して、包括的な戦後計画を顧みる余裕がないという空気が支配的であった。官僚会議でも閣僚会議でも『ベヴァリッジ報告』の棚上げが主流となり、その代わりとして、ベヴァリッジ報告が前提とする数々の仮定（家族手当、包括的保健・リハビリサービス、完全雇用）や数字（平均失業率が 10%）を吟味するために、国民所得の綿密な推計がここでも急務とされたのである。

他方、ベヴァリッジは報告書を完成させる直前から、次の目標を「無為からの解放」（つまり高度な雇用の維持）に定めており、オックスフォード大学を拠点に民間人の専門家と――政府のベヴァリッジ接触禁止令も出されたため――議論を重ねることになった。協力者はバーバラ・ウットン、ジョーン・ロビンソン、ニコラス・カルドア、シューマッハ等、ケインズに強く影響を受けた若手経済学者であった。Harris (1997: 434)の指摘によれば、シューマッハに説得されたベヴァリッジは 1943 年９月にメモ[67]を提出して、初めてケインズ的な

---

[65] Keynes (1979, vol. 22: 218), 'Notes on the Budget I', 21 September 1940.
[66] Keynes (1980, vo. 27: 354), a letter to W. H. Beveridge, 17 March 1942.
[67] Nuffield College Private Conferences, 1941-1955, Archive Collections, the Library of

考え[68]を表明した。ここでケインズ的とは、どんなに組織化されても労働需要は常に過小になるので、主要な「社会化された」産業において、「社会化された」需要に基づいた公共投資・民間投資によって完全雇用を支えるべきだ、とする考えである。ケインズの「投資のやや広範な社会化」（Keynes 1936: 378）と同様に、ベヴァリッジも「社会化」を重視していること[69]がわかる。

なお、ベヴァリッジが題名で明示したように、「自由社会における」という限定は非常に重要である。完全雇用の実現は喫緊の課題であるが、すべてを犠牲にして達成するものでもない。全体主義国家とは異なり、あくまで「自由社会」という前提が必要である。そこで、この言葉の内容が吟味される。自由社会とは「市民の基本的自由権がすべて確保されているという条件に従っていること」（Beveridge 1944: para.11）である。そして基本的自由権とは、信仰・言論・執筆・研究・教育という不可侵な権利を指す。

その他の細目として、政治その他を目的とする集会・結社の自由、職業選択の自由、個人所得を処分する自由が問題となる。ベヴァリッジは次のような問を発する。完全雇用では賃金や物価が累積的に上昇する可能性があり、このときにストライキ権を付与された団体交渉は絶対的な自由として擁護されるべきだろうか。職業選択の自由が進みすぎると、離業によって常に生産性を高めることが難しいのではないか。個人の貯蓄と消費の比を自由に決められる場合、労働需要と労働による生産物が必ずしも供給に一致しない可能性も出てくるのではないか（Beveridge 1944: para.13, 14, 15）。全体主義の政策にならないように留意しながら、基本的自由権を堅持しつつ、こうした困難をどのように解決していくかがベヴァリッジの関心事であった。そのためには、基本的自由権の周辺にある自由権をいくらか制限しなければならないという判断であった。

ベヴァリッジはこのような私的報告書を公表しようとしていたが、その人気に危機感を募らせたのが政府首脳であった。保守党と労働党が同居する挙国一

Nuffield College, Oxford. “Full Productive Employment in a Free Society”, by W. H. Beveridge, 1-11, 8 September 1943.　この文書の閲覧について、上宮智之氏（大阪経済大学）の支援を受けた。

68 「保障される雇用は生産的で進歩的でなくてはならない。単なる時間の無駄（穴を掘って埋める）や破壊的（戦争、軍備、ナチス的方法）であるような職業によって、問題を解決することは排除している」（３段落目）。ここでは、ケインズ的な考えに向けられた批判をあらかじめ回避している。

69 報告書本体でも「有効需要の社会化」（Beveridge 1944: para.271）が推奨された。具体的には国家投資局を設立し、情報分析を行い、官民投資の支援や制御を行う権限が与えられる（Beveridge 1944: para.241）。

致内閣は、閣内不一致を恐れ、戦争の遂行に集中したいため、雇用問題に関する政府の公式見解を早急に公表する必要に迫られた。公表のスピードこそ問題（Booth 1989: 113）であり、ベヴァリッジの私的報告書を出し抜かなくてはいけなかった。既にその報告書は 1944 年 5 月に印刷所に送られていたが、民間の出版物は後回しにされたため、政府は同月に白書を公表して、何とか面目を保った。社会保険に関する白書[70]も、『ベヴァリッジ報告』が持つ基本的理念のいくつかを大きく逸脱する形で、1944 年 9 月に公表された。結局、『自由社会における完全雇用』が出版されたのは 1944 年 11 月であった。

労働党も独自に戦後雇用の問題に取り組んでいた。内務大臣モリソンは『ベヴァリッジ報告』に刺激を受け、この報告書に冷淡な政府内の空気に対抗すべく、商務大臣ドールトンに労働党の内部でこの問題を議論するように頼んだ（Toye 2003: 146）。ピグーやケインズに学んだ財政学者でもあったドールトンは党大会のための発表原稿を自ら仕上げ、1944 年 4 月に公表した。ドールトンはこの報告書がだいたいにおいてケインズ的であることを認めた。つまり、完全雇用を維持するには、購買力を維持しなければならないし、総支出と貯蓄を一致させなくてはいけない。ただし、社会主義的な部分――銀行や鉄鋼など、主要産業の活動や配置を政府主導に任せよ――も残していた。ドールトンは白書のレースに勝利したことを誇った。政府の白書やベヴァリッジの私的報告書より早く出版され、内容も最も濃密（最小の文字数）だったのである（Dalton 1957: 422-423）。

以上のように、政府側も労働党側も、『ベヴァリッジ報告』とその周辺に直接的・間接的に影響されてケインズ的な考えを吸収して、完全雇用問題をそれぞれの立場で深めていった。このような時代背景を前提に、『自由社会』のテキストマイニングに移ろう。

4-2　頻度分析による比較

本節では実際に KH Coder[71]（Ver. 2.00f; 2015.12.29）というソフトウェアを用いた。KH Coder では、主に 2 つの段階からなる分析手順を想定して設計が行われている。まず段階 1 では、データ中から語を自動的に取り出して、その

---

70 *Social Insurance, Part I*, Presented by the Minister of Reconstruction to Parliament by Command of His Majesty, Cmd 6550, September 1944.

71 このソフトは樋口耕一氏（立命館大学）が開発したフリーウェアである。主な機能と分析手順は http://khc.sourceforge.net/diagram.html にある。

結果を集計・解析する。これによって、分析者の予断をなるべく交えずに、データの特徴を探り、テキストを要約することが可能となる。次に段階２では、分析者が「こういう表現があれば、コンセプト A が出現していたと見なす」といった指定（コーディングルール作成）を積極的かつ明示的に行い、データ中からコンセプトを取り出し、その結果を集計・解析することで、分析を深めることができる。なお、第３節で紹介した TinyTextMiner とあえて異なるソフトを用いて、使いやすさの比較も考察する。

最初にテキストの前処理を行った[72]。1944 年版オリジナルの文書について、注・補遺を省いた本文のみを対象にした。第 1 部からあとがきまで 8 カ所に部が分かれており、この部の見出しに<h1>1 INTRODUCTION AND SUMMARY </h1>というタグのマークを施した。このマーキングは、対象テキストを　１．文（ピリオドで認識）、２．段落（一行空けで認識）、３．章立て（h1 というタグで認識）のいずれかで集計するための工夫である。<h2>introduction</h2> のように、章の下に節を挟み込むこともできる。ここでは部 part という単位のみでタグ付けした。ソフトがテキストをうまく処理できたかどうかを「前処理」→「分析対象ファイルのチェック」　を実行することで、自動的に補正することができる。対象テキストは　文：3880、段落：661、部（h1）：９　という具合に認識された。

まず基本となる頻度分析である。「ツール」→「抽出語」→「抽出語リスト」によって、名詞・固有名詞・代名詞・形容詞・形容動詞・動詞などが自動的に検出される。ここで、頭文字または単語すべてが大文字に表記されている場合を「固有名詞」であると自動的に判定しているようだ。

名詞の頻度分析から始める。頻出上位 20 位までを「補正前」とした。いったんこの表を作成した後、改めて本文を精査すると、小見出しの大部分がすべて大文字として表記されていた。そこで、一般的な単語に関しては大文字を無視して固有名詞ではなく、目視によって名詞に分類し直した（表 4-1）。それが「補正後」である。この補正を施しても、上位 20 位の中で変動はほとんどない。employment が１位、unemployment が２位、industry が３位、demand が４位、war が５位となっている。

[72] KH Coder を起動し、「プロジェクト」→「新規」によって対象ファイルを指定する。次に、「前処理」→「前処理の実行」を選択する。

Noun	補正前	ProperNoun		Noun	補正後
employment	604	50	1	employment	654
unemployment	565	2	2	unemployment	567
industry	422	3	3	industry	425
demand	404	2	4	demand	406
war	353	26	5	labor	374
labor	348		6	war	353
policy	338	14	7	policy	352
country	295	4	8	country	299
outlay	259	1	9	outlay	260
man	237	12	10	trade	247
trade	235		11	man	237
year	222		12	year	222
investment	191	16	13	investment	207
rate	179		14	rate	179
cent	164		15	cent	164
problem	163	1	16	problem	164
time	152	2	17	time	154
state	148		18	state	148
condition	143	5	19	condition	148
fluctuation	138		20	fluctuation	138

表 4-1　頻出語 20 位まで（名詞）

表 4-1 から表 4-3 を比べれば、『一般理論』本体との差も判明する。両著ともに上位に入っているのは、employment, demand, labor（実際には labour、ソフトの都合でアメリカ語綴りに自動的に変換されている）, investment, rate の５つのみである。ただし、いずれも非常に重要な単語となっている。『一般理論』のみに頻出する単語は、interest, money, (change), capital, income, (cost), (output), (price), (quality), (term), level, consumption, (increase), (value), (efficiency)の 15 個である。そのうち、(　　)の付いた単語は『一般理論』の書

評にも頻出しないもの[73]（９個）である。つまり、残りの６単語は書評では見逃していないものの、『自由社会』では頻度としては脱落した概念となっている。この６つのうち、interest, money, capital は貨幣概念に強く関わるもの、income, consumption は投資以外の総需要項目となっている。

逆に、『自由社会』のみに頻出となる単語もある。unemployment, industry, war, policy, country, outlay, man, trade, year, cent, problem, state, condition, fluctuation の 14 個である（time は『一般理論』で 26 位に顔を出すので、ここでは除いた）。いずれもベヴァリッジの時代や理念を反映した単語と捉えることもできる。中でも上位５位までに入っている unemployment, industry, war の３つは最重要である。さらに、７位の policy や９位の outlay の独自性もわかる（outlay は income や consumption の代替となる用語と見なせる）。

Adj	補正前	ProperNoun		Adj	補正後
full	389	30	1	full	419
other	307		2	other	307
private	222	3	3	private	225
first	183	6	4	first	189
public	158	5	5	public	163
more	145	23	6	international	156
economic	141	50	7	social	151
total	140	51	8	national	150
general	137	7	9	economic	148
international	133		10	more	145
such	131	8	11	general	145
new	120	2	12	total	142
possible	116		13	such	131
unemployed	112	10	14	new	130
same	111		15	possible	116
social	101	4	16	unemployed	116
national	99		17	same	111
industrial	97	9	18	industrial	106

[73] 下平・小峯・松山（2012: 17）の表１を見よ。

many	97		19	many	97
particular	95		20	particular	95

表 4-2　頻出語 20 位まで（形容詞）

rate	709	marginal	342
interest	684	other	251
employment	597	real	219
investment	526	same	192
money	457	aggregate	160
change	394	full	150
capital	392	current	148
income 385	385	new	146
cost	304	such	146
demand	294	equal	135
output	274	more	128
price	151	effective	126
quantity	249	economic	113
term	231	different	109
level	220	classical	102
consumption	219	general	102
increase	216	certain	94
labor	213	due	94
value	210	net	93
efficiency	202	prospective	82

表 4-3　『一般理論』の頻度分析（名詞と形容詞）

次に形容詞の頻度を調べる（表 4-2）。名詞と同様に、補正を施した。つまり、固有名詞（大文字の単語）に分類されている形容詞を、目視によって形容詞に分類し直したのである（元の 20 位の範囲のみ）。名詞の場合と異なり、大きく順位が変動した。つまり social, national という二語は見出し語としてよく使われていたため、補正前は順位が低かった（16 位と 17 位）が、補正後はこの二語と economic という形容詞がほぼ同一頻度となった（７位、８位、９位）。ま

たこの上に international が 6 位として入っている。他方、『一般理論』で特徴的なのは、marginal, aggregate, effective, classical である。いずれも次に来る単語が予想できるほど、重要で馴染み深い専門語であることがわかる。

ここで（投資や有効需要の）「社会化」という特殊な用語がベヴァリッジとケインズで共有されていることが、テキストマイニングの前から判明していた。そこで social という用語に関連する単語を集めてみた（表 4-4）。このように上位頻出語だけでなく、ピンポイントである関連語を集めてみると、案外にも全体としては多くの使用回数となっていることもわかる。１つの解釈として、ベヴァリッジは経済問題を《社会》あるいは《社会化》という観点から――または《社会主義》との関連で――述べていた、と判断される。

抽出語	頻度	品詞
social	101	形容詞
SOCIAL	50	固有名詞
society	43	名詞
socialism	10	名詞
socialization	10	名詞
socialize	4	動詞
socially	3	形容動詞
socialist	3	形容詞
anti-social	1	形容詞
socialized	1	形容詞

表 4-4　social に関係する単語

固有名詞は（Britain が上位として目立つものの）人名だけに注目した（表 4-5）。ケインズが１位、カルドアが２位である。その他、ピグー、ベヴァリッジ、ベヴィン、クレイ[74]が複数回に渡って言及されている。よく知られたように、カルドアはベヴァリッジが組織した私的調査委員会の重要メンバーだっただけでなく、『自由社会』の付録Ｃ「イギリスにおける完全雇用問題の量的側面」を署

[74] 当初、目視によるカウントで、Clay が人名であることを見逃していた。その後に原典の索引を見たところ、Clay の見落としに気づいた。このように、マイニングの結果と原典とを常に行き来しなければならない。

名付きで書いている。ハイエクは根拠を挙げることなしに、この事情を次のように述べている。

> 「彼はベヴァリッジの雇用に関する本を書いた。これはよく知られた事実である。そこにある経済学はすべてカルドアのものだ。このような本を書く能力はまったくなかったはずだ。あの本にはカルドアと認められる章が入っているが、全体が彼によるものだ。あの章はベヴァリッジが関わりたくないと思った部分だが、それを理解できなかったからである。」（Hayek 1994: 86／訳 84）

ロビンズも同様に、この本は「助言に基づいて書かれたと悪評高かった」（Robbins 1971: 136／訳 147）と回顧した。有力な経済学者によるこの印象によって、カルドアに全面依拠したベヴァリッジという評価が形成された。

しかし、Harris (1997: 432-443)は、技術的な議論についてベヴァリッジは限定的な役割しか担っていないが、統合と詳説を成し遂げ、戦後の社会保障という広い文脈に完全雇用の問題を位置づけた、と総合的に評価した。この研究は本書の成立過程を草稿やメモに基づいて厳密に検証した決定版である。なお、ベヴァリッジがケインズ的考えに転換した重大な転機は、4-1 で前述したように、カルドアというよりもシューマッハとの討論だった。

Keynes	24
Kaldor	15
Pigou	5
Beveridge	3
Bevin	2
Clay	2

表 4-5　人名

以上の事情を鑑みて、さらに本書の索引と対照してみた場合、索引ではカルドアは一カ所しかページ数の指示がない。実際の言及は 15 回である。索引から受ける印象よりも、多くの言及がある。シューマッハは索引にもなく、頻度分析にも引っかからなかった。この結果だけを見ると、カルドアを重視するハイ

エクやロビンズの言説が正しそうであるが、草稿を詳細に分析した先行研究によれば、彼らの評価は正しくない。この事例から、本文の頻度分析のみに頼るとテキストを大きく誤読する可能性も指摘できる。

4-3　クラスター分析の有用性

次にクラスター分析を試みる。下平・小峯・松山（2012）ではあまり有効な結果が出なかった分析だが、書評という文書ごとではなく、段落ごとのクラスター（房）を作ることによって、この分析の有用性を確かめたい。この分析は個々のデータの類似度をある種の《距離》として表現し、距離の近いデータ同士をまとめて、テキスト全体をいくつかにグループ化する分析方法である（石田・金編 2012: 11）。また、分類基準が何もないときに、データが持つ情報に基づいてグループ分けができる（田崎編 2015: 175）。距離として表現する類似性には様々な種類があり、データの特性に応じて取捨選択するしかない。テキストマイニングでは、通常の「ユークリッド距離」よりは Jaccard 係数を利用するのが良い（石田・金編 2012: 12；樋口 2014: 152）。選択した距離の種類によって分析結果が大きく変わることもありえるため、クラスター分析では唯一の解を求める最終手段というよりは、データの意味づけを模索するための便宜的手段と捉えた方が良い（石田・小林 2013: 112）。

テキストマイニングに必要な距離の概念に若干、触れておこう。共起頻度、Jaccard 係数、Simposon 係数を次のように定義する（松尾ほか 2005: 48）。

$$\text{共起頻度} = |X \cap Y|$$
$$\text{Jaccard 係数} = |X \cap Y| / |X \cup Y|$$
$$\text{Simpson 係数} = |X \cap Y| / \min(|X|, |Y|)$$

ただし、X と Y はある集合における要素の数である。共起頻度は言わば共起の絶対値である。残りの２つはこの絶対値の相対的な位置を確かめているため、０から１の値を取る。Jaccard 係数では分母を全要素の数としているのに対して、Simpson 係数では２つの集合のうち小さな要素の数を分母としている。Jaccard 係数[75]は質的データの解析によく用いられるが、２つの集合で要素数に大きな隔

[75] 樋口（2014: 152）は、１つの文書に含まれる単語の数が少なく、それぞれの単語が一部の文書中にしか含まれていないようなデータの場合に Jaccard 係数が相応しいと判断している。この係数は単語が共起しているかどうかを重視している。石田・金編（2012: 12）で

たりがある場合、係数が非常に小さくなるという欠点も持つ。Simpson 係数ではこの欠点が克服されているが、一方が非常に少ない要素の場合は、係数が 1 に近づきすぎる（石田・金編 2012: 12）。このうち、クラスター分析では項目間の距離（非類似性）を用いるため、Jaccard 係数を 1 から引いて、類似しているほど値が小さくなるような距離を定義する（鈴木 2017: 113）。

$$\text{Jaccard 距離} = 1 - |X \cap Y| / |X \cup Y|$$

KH Coder ではこの Jaccard 距離を用いている。

さらにクラスターを作成するための方法として、ウォード法 Ward's Method が頻繁に用いられる。これはクラスター内の平方和を最小にするように、クラスターを併合していく方法である。つまり分散の情報を用いており、散らばり具合が併合後もあまり増加しないならば、似たもの同士がまとまったと判断する。重心法など他の方法に比べて、ウォード法では鎖効果――併合されるクラスターが 1 つずつ順番に増えていくため、類似のクラスターが複数出ないこと――が起こりにくいという利点がある（豊田編 2008: 190）。

クラスター分析には階層的と非階層的がある。前者では、最も近い関係のグループがまず小さな階層を作り、その階層を含む形で別のグループとの階層が定まり、最後にはデータ全体を包括するクラスターが階層的に形成されることになる。この様子を可視化したものが樹形図 dendrogram である。この図はある程度、限られたデータの場合に図示しやすい。様々な定義がある距離のうちどれを採用するかによって、大きく樹形図の形状が変化する。また、図が非常に大きくなってしまうので、空間の節約という点では分が悪い（樋口 2014: 38）。また、異なった方法による結果がしばしば大きく異なる（石田・金編 2012: 12）。あくまで、データ同士がどのように結びつくかの一例として、発想法を豊かにする手段と考えたい。

他方、非階層的クラスター分析は、あらかじめ指定したクラスター数で観察対象を分類する。その利点は膨大なデータの場合、クラスター数[76]の目星（だいたい 2 から 7）を付けてから、その前後の数を実施して結果を比較した方が、手間がかからない点にある（豊田編 2008: 193）。非階層的クラスター分析の

はデータが質的な場合、または 0, 1 である場合にこの係数が用いられるとする。

[76] 結合距離が長くなりそうな部分で、クラスター数を決定するという方法もある（田崎編 2015: 183）。

場合、クラスターが設定された数だけ描かれるだけで、階層的分析のように、最後に唯一の枝が描かれて樹形図が完結することはない。

以上の意義と限界を前提に、データの大きさも考え、『自由社会』の階層的クラスター分析を試みた（図 4-1）。集計単位は「段落」として、最小出現数を 105 とした。品詞は名詞と形容詞のみで、固有名詞は省いている。Ward 法と Jaccard 係数を選択し、クラスター数は Auto とした[77]。単語に一番近い場所でコの字型にまとめられていれば、その単語が最も近い関係にあることを示している。単語の左側にある棒グラフは出現頻度数である。さて、一番近い単語は full と employment であることが図から判明する。二番目は labour と demand である。三番目は private と public である。四番目が international と trade である。五番目が war と peace である。一番目から三番目まではごく自然な結びつきだが、四番目と五番目は、従来の研究では強調されていなかった。以下、ほぼ数値が変わらない形で、（labour と demand）と結びついた industry、rate と cent、man と job、（private と public）と結びついた investment、outlay と state、(international と trade) と結びついた country、(full と employment) と結びついた policy などが続く。

テキスト全体を 10 のクラスターに分けた場合、図 4-1 のような樹形図になった。図の中の破線はクラスターの切断位置を示しており、使用された語数の平方根を四捨五入した数[78]に対応している（田崎編 2015: 206）。このように導出されたクラスターを上から順番に括ってみると、次のように分類できる。

77 集計単位を「段落」とした理由は、文と部の中間単位だからである。「最小出現数」は任意に選べる。あまり小さな数だと該当する単語が多すぎる。なお、今回は固有名詞を省いているので、見出し部分（大文字）は考慮しないことになる。

78 クラスター数を「Auto」にした場合は 7 となったが、多くの単語が同じクラスターに入っていた。そこでもう少し分割するために、「調整」ボタンを押して 10 を割り当てた。

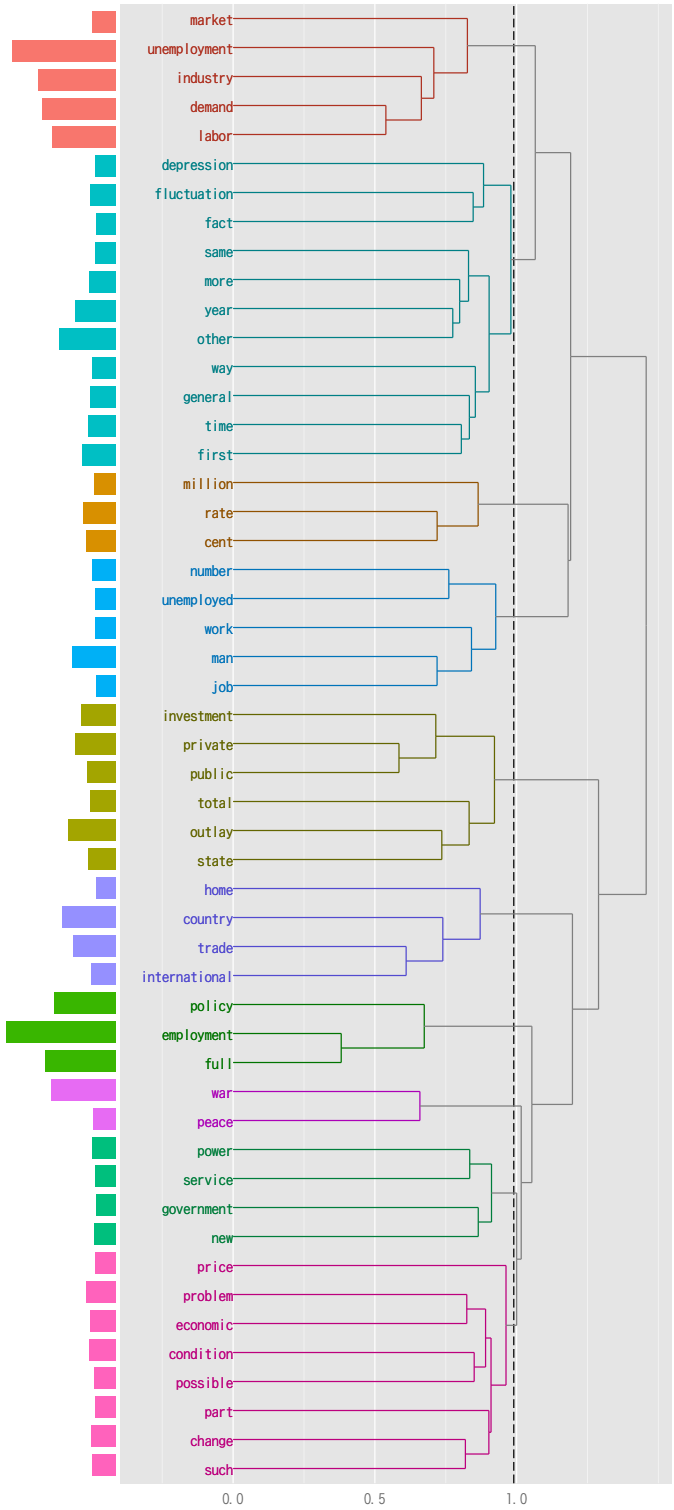


図 4-1　階層的クラスター分析（10 房）

第１クラスター：失業・労働需要・産業に関する話題（market, unemployment industry, demand, labour）
第２クラスター：不況・変動・時間に関する話題（depression, fluctuation, fact, same, more, year, other, way, general, time, first）
第３クラスター：単位（数や%）に関する話題（million, rate, cent）
第４クラスター：失業者数や仕事に関する話題（number, unemployed, work, man, job）
第５クラスター：民間投資・公共投資・総支出に関する話題（investment, private, public, total, outlay, state）
第６クラスター：国際貿易に関する話題（home, country, trade, international）
第７クラスター：完全雇用政策に関する話題（policy, employment, full）
第８クラスター：戦争と平和に関する話題（war, peace）
第９クラスター：政府の新しいサービスに関する話題（power, service, government, new）
第 10 クラスター：経済問題・可変的な条件・価格に関する話題（price, problem, economic, condition, possible, part, change, such）

こうしたクラスターは『自由社会』における重大な論点を適切にまとめていると判断できる。

4-4　共起ネットワーク分析と中心性

次に共起ネットワーク分析[79]に移る。共起 collocation, co-occurrence とは、特に自然言語処理において、ある指定された範囲のテキスト（文書全体・段落・文・前後 n 語など）の中で、ある文字列・単語と別の文字列・単語が同時に出現することである。n-gram と異なり、検索語と共起語が隣接している必要はない（小林 2017a: 130）。共起頻度を示すには、集計単位・最小出現数・対象品詞などを確定しなければならない（小林 2017b: 113）。この確定に関しては一般的な法則はなく、試行錯誤を重ねる必要がある。

[79] N-gram 分析とも似るが、村井編（2014: 16-17）によれば、非連続的で緩やかな関係性を分析したい場合に適する。

この共起の構造（ネットワーク）を視覚的に表現する代表的な方法として、共起ネットワーク分析がある。ネットワーク分析でもあるので、数学のグラフ理論を応用することになる。グラフ理論で頂点 vertex に当たるものを各単語、辺（直線）edge に当たるものを共起関係（単語と単語の近さ、距離）と解釈して、共起が強い単語同士を線で結ぶ。線分の太さを共起の強さと表現することもできる。KH Coder では結びつきの度合いを色分けすることもできる。また単語の頻出度合いに応じて、各単語の円の大きさを変えて描くこともできる。ここでは共起関係の強弱[80]を Jaccard 係数によって計算している。なお、低い頻度だが強い結びつきの共起を重視する際は、「相互情報量」[81]という指標を使うこともできる（小林 2017b: 117-118）。

この分析によって描かれた図を解釈する際に、留意すべき点が２つある。１つは単語と単語の距離に意味はないことである（樋口 2014: 155）。図を書く都合で自動的に配置されるので、問題は語と語の物理的な近さではなく、その間に線分が引かれているか（さらにその線分の太さがどうか）である。もう１つは結果の解釈を行う際に、必ず元のテキストに戻って文意や文脈を確認しなければならないことである（石田・金編 2012: 123）。この確認には「KWIC コンコーダンス」というコマンドが便利である。

共起ネットワーク分析の視覚化には、中心性の分析とサブグラフ検出に分けることができる。前者は、それぞれの単語がネットワーク構造の中で、どの程度中心的な役割を果たしているかを示す。なおネットワーク分析では、様々な「中心性」の概念[82]があり、KH Coder でも次の三種類を自動的に検出することができる。

①媒介中心性 betweenness centrality：頂点同士の中間に位置する、言わば情報の媒介地点（交通の要所）。各要素を最短経路で結んだ場合に、ある経路を要素が追加する回数の多さにも喩えられる。人間関係に置き換えると、「事情通」であり、全体に大きな影響を与える。

②次数中心性 degree centrality：他の要素への結びつきの多さを示す。ネットワークのハブに当たり、各要素が共通して持つもの。各頂点に接続

---

80 ソフトの開発者は、無理矢理に目安を設けるならば、Jaccard 係数の 0.1 を弱い連関、0.3 を強い連関と見なしている。KH Coder 掲示板（2018.8.28 閲覧）
http://koichi.nihon.to/cgi-bin/bbs_khn/khcf.cgi?no=122&mode=allread

81 $\log_2$ (共起頻度×データの総語数)/(検索語の単純頻度×共起語の単純頻度)。

82 田中（2014: 42-43）、（独）情報処理推進機構編（2017: 6）、鈴木（2017: 67）の表現を参考にした。

する辺（線分）の数のみを抽出している。すべての辺が同等の価値を持つと仮定。人間関係に置き換えると、「人気者」である。

③固有ベクトル eigenvector centrality : ある頂点の中心性を評価する場合に、その頂点と隣接する頂点の中心性を反映させる。ネットワーク全体の構造を反映しているもの[83]。１つのリンクは重要度の高いもの、もう１つのリンクは重要度の低いものという２つのリンクを持つ頂点と、どちらも重要度が低い頂点では、前者を高く評価する（次数中心性では、２という数値により両者は同等）。人間関係に置き換えると、「大事な場面に常にいる人」である。

後者の視覚化はサブグラフ検出である。これは比較的強く互いに結びついている単語を自動的に検出し、グループ分けの結果を色分けして示すものである。KH Coder では３種類の分析が可能であり、それぞれの計算方法によって、グループの数も異なってくる。

２つの視覚化を使って、『一般理論』との比較を意識しながら、『自由社会』の分析を進めよう（図 4-2 と図 4-3）。「ツール」→「抽出語」→「共起ネットワーク」で実行できる[84]。中心性分析によれば、『自由社会』で中心性（媒介）を持つ単語は、employment, unemployment, private の３つである。最初の２つは論じられている題材を鑑みれば当然であるが、３番目はどう解釈すべきだろうか。ここで中心性分析が析出するのは、必ずしも最重要概念の単語という意味ではなく、他の言葉と繋がりが強い単語を含んでいることである。そのため、「サブグラフ検出」によれば、自らのグループのみならず、別々の結びつきに分類された full（または policy）や employment とも近い距離にあるという特徴が窺える。言わば、「完全雇用政策」という単語と近いと判定された。１つの解釈として、この政策は民間と大いに関係がある、というメッセージとも捉えられる。実際、ベヴァリッジは「産業の大部分は私的なリスクを負う私的な企業によって担われ続ける」（Beveridge 1944: 37）と断言し、完全雇用政策と私的企業の両立性を強調しているのである（Beveridge 1944: 205-207）。

---

83 隣接している頂点はさらにその隣接する頂点の中心性を反映しているという意味で、この中心性は到達可能なすべての頂点の中心性を映し出している。ただし、そのネットワークがいくつかのサブグラフに完全に分離している場合は、その有効性が減じる。鈴木（2017: 61）を参照のこと。

84 今回も集計単位は「段落」、最小出現数は 105、品詞は「名詞」と「形容詞」とした。中心性もサブグラフ検出も「媒介」を選択した。

この意味で、テキストマイニングによる分類の妥当性が、実際の文献熟読によって確認されたことになる。

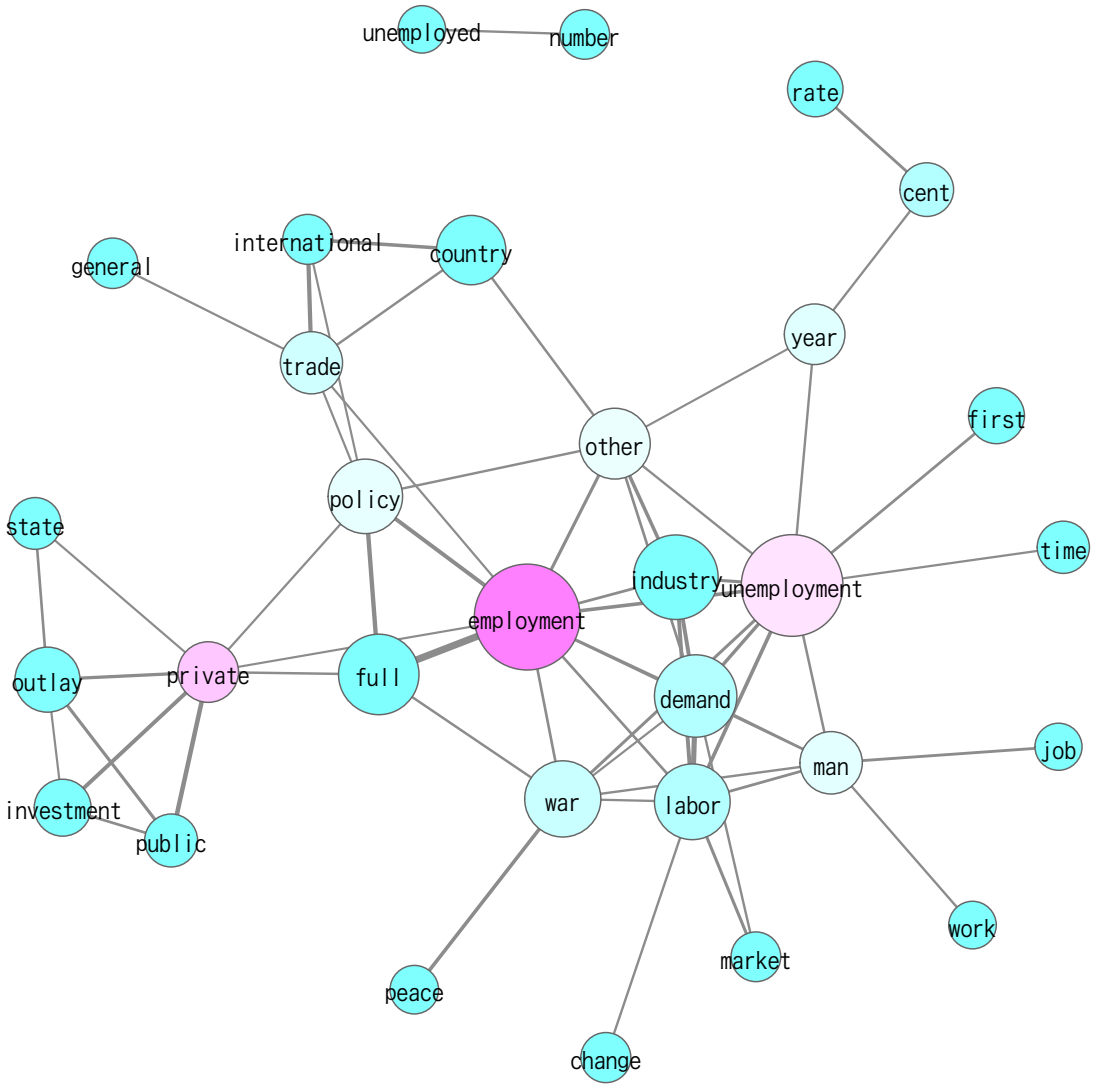


図 4-2　共起ネットワーク分析（中心性）：『自由社会』

他方、『一般理論』では中心性のある単語は employment しかない。他の単語と一番深い関わりとなる。題名の最初に来ることからわかるように、雇用はどのように決定されるのかという理論的関心が前面に出てきており、この結果に矛盾はない。「サブグラフ検出」からわかるのは、『自由社会』でも『一般理論』でも 6 つのブロックに分類されていることである（図 4-4 と図 4-5）。し

かし、その中身は大きく違う。前者では、支出、国際貿易、失業雇用と労働需要、失業者数、単位という分かれ方だが、後者では、投資と限界効率と雇用、

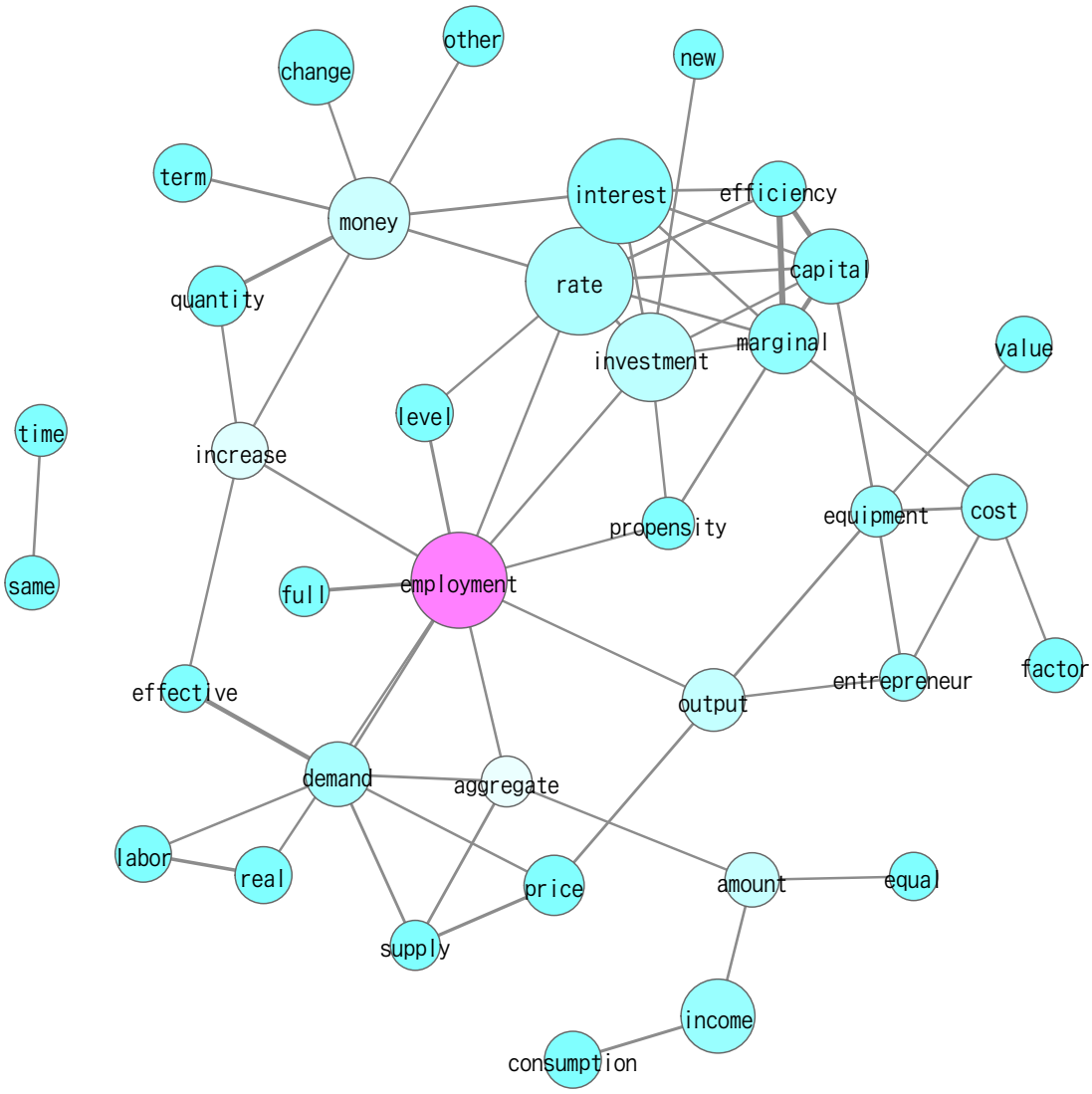


図 4-3　共起ネットワーク分析（中心性）：『一般理論』

貨幣量、有効需要・総供給・労働、所得と雇用、生産の費用と設備、時間という分かれ方である。これだけでも『自由社会』がより政策指向であり、『一般理論』がより理論指向であることがこの解析からも確認できた。

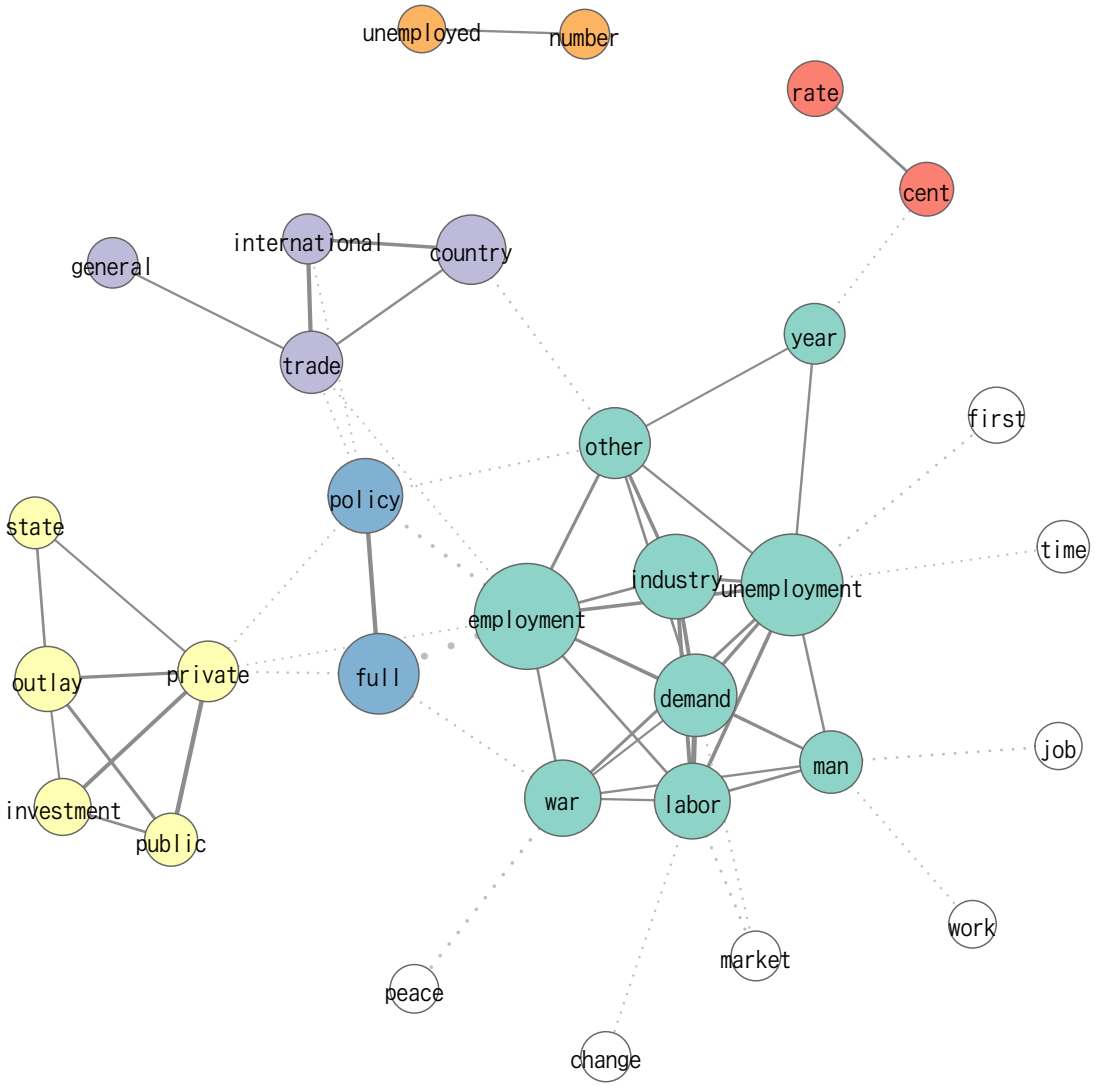


図 4-4　共起ネットワーク分析（サブグラフ検出）：『自由社会』

4-5　特徴語分析とコロケーション統計

この項では「特徴語分析」を行う。テキスト全体とテキストの一部（例：各章ごと、本文と補遺）の差・特徴を掴みたいときに便利である。ここでも Jaccard 係数による類似性を出現頻度と共に測定し、その上位を各章の「特徴語」と定義した（樋口 2014: 38-39）。この概念は単なる絶対的な頻度だけでなく、相対的な頻度も考慮に入れていることになる[85]。「ツール」→「外部変数と見出し」

[85] ソフトの開発者は「特徴語」の定義を「それぞれの部において、特に多く出現している言葉」（樋口 2014: 38）としているが、この定義では不正確と思われる。

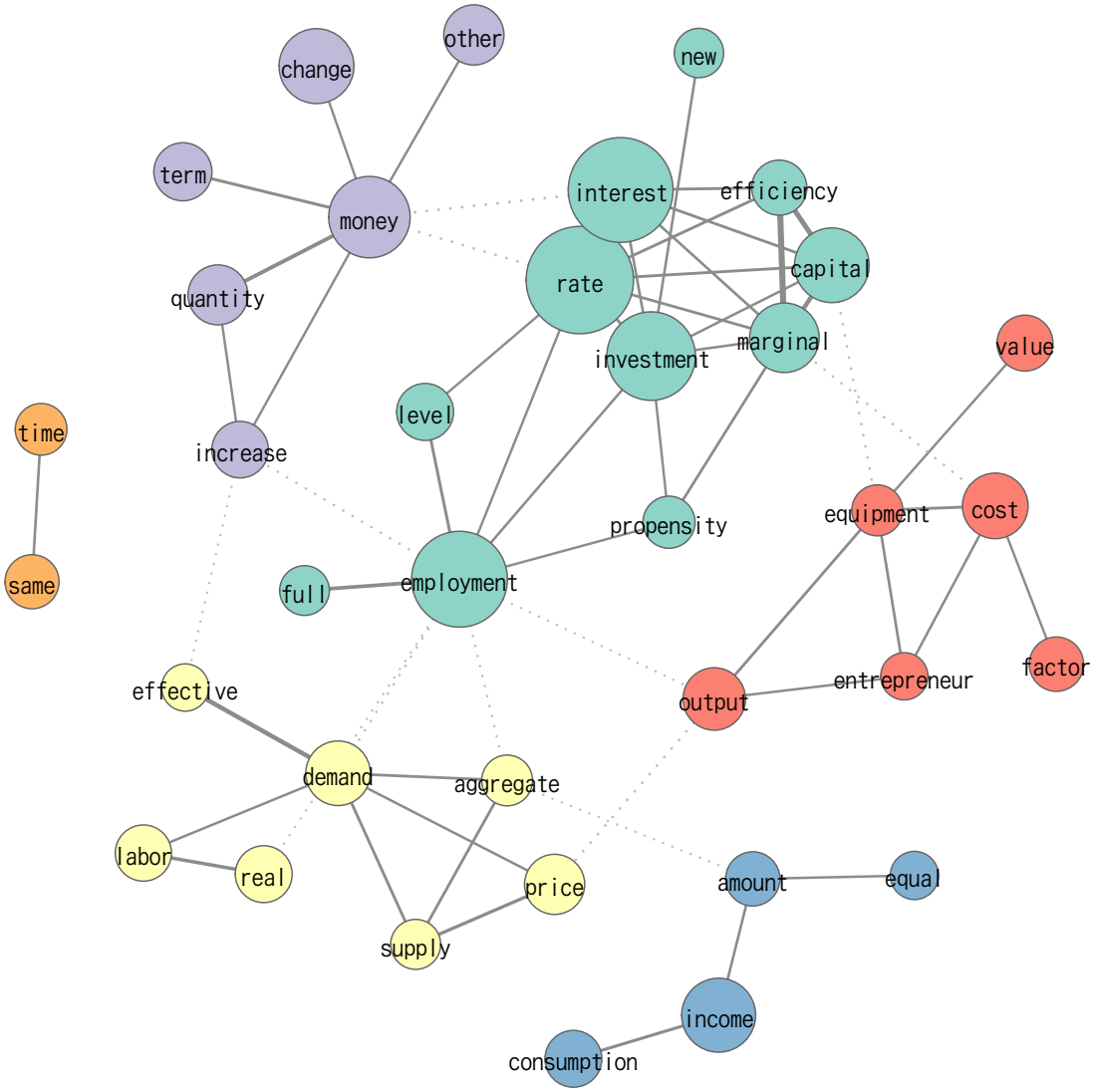


図 4-5　共起ネットワーク分析（サブグラフ検出）：『一般理論』

→「リスト」を選び、析出したい章（部）を選択したら、「特徴語」をクリックする。その際、単位を「文」にしておく。考慮する範囲が各単元という小さな単位になっているので、計算する単位も「段落」ではなく「文」にする。

『自由社会』のテキスト全体は、各部 part に従って、次の 9 つに分かれている（章や節には分かれていない）。

序文
第 1 部　導入と要約

第２部　平和における失業
第３部　戦争における失業
第４部　平和に向けた完全雇用政策
第５部　完全雇用の国内的含意
第６部　完全雇用の国際的含意
第７部　完全雇用と社会的良心
あとがき

表 4-6 と表 4-7 からは次のような特徴が窺える。序文と第１部には明確な特徴語は見いだしにくい。第２部では unemployment と industry が上位にあった。第３部 war と peace が突出している。第４部では outlay, employment, full が上位にある。なお outlay は全テキスト 192 回で頻度だが、第４部では 148 回という具合に相対頻度では突出していた。この傾向は taxation（同 67 回のうち、53 回）にも見られる。第５部では price, wages など価格関連の単語が上位にある。第６部では貿易に関連した用語が多数を占める。第７部では life, war, man, work など生活に関連した用語の後に、liberty や free など自由に関連した用語が続いている。あとがきは政府の白書『雇用政策』をベヴァリッジが読了した後、その感想を付け加えた部分であり、government, policy, diagnosis などの単

A	B	C	D	E	F
	Part2				
	抽出語	品詞	全体	共起	Jaccard係数
1	unemployment	Noun	479 (0.122)	221 (0.257)	0.1977
2	industry	Noun	335 (0.086)	166 (0.193)	0.1613
3	labor	Noun	281 (0.072)	122 (0.142)	0.1197
4	rate	Noun	156 (0.040)	93 (0.108)	0.1008
5	demand	Noun	337 (0.086)	108 (0.126)	0.0992
6	year	Noun	172 (0.044)	82 (0.095)	0.0863
7	other	Adj	294 (0.075)	75 (0.087)	0.0695
8	number	Noun	101 (0.026)	57 (0.066)	0.0631
9	first	Adj	179 (0.046)	58 (0.067)	0.0591
10	fact	Noun	99 (0.025)	53 (0.062)	0.0585

表 4-6　特徴語（第２部）

A	B	C	D	E	F	G
	Part4					
	抽出語	品詞	全体	共起	Jaccard係数	
1	outlay	Noun	192 (0.049)	148 (0.131)	0.1265	
2	employment	Noun	551 (0.141)	187 (0.166)	0.1255	
3	full	Adj	368 (0.094)	147 (0.131)	0.1091	
4	private	Adj	195 (0.050)	108 (0.096)	0.089	
5	investment	Noun	159 (0.041)	91 (0.081)	0.0762	
6	public	Adj	136 (0.035)	86 (0.076)	0.0731	
7	policy	Noun	294 (0.075)	86 (0.076)	0.0645	
8	state	Noun	137 (0.035)	69 (0.061)	0.0578	
9	taxation	Noun	67 (0.017)	53 (0.047)	0.0465	
10	national	Adj	92 (0.023)	50 (0.044)	0.0428	

表 4-7　特徴語（第 4 部）

語が抽出されている。以上の順位は Jaccard 係数に基づいているが、微少な数字の差によって順位が定まっており、実際にどの程度、順位が重要なのかはさほど自明ではない。

むしろ、こうした特徴語の分析を進めていく際に、抽出された用語が実際にはどのような文脈で用いられているかを確認する作業が絶対的に必要である。KH Coder では「KWIC[86]コンコーダンス」というコマンドがあるので、特定の品詞や特定の活用形を指定した後に、実行できる。表 4-7 が実際の画面である。コンコーダンス画面にまずは一行分の文が表示されるが、これよりも広い範囲で文脈を確認したい場合は、全体の文書を表示することもできる。

こうしたコンコーダンス検索を行っても、非常に多数の検索結果が得られた場合は、その１つ１つの文脈をすべて辿ることは事実上、不可能である。そこである特徴語が使われている文脈を総体として読み取りたい場合に、「コロケーション統計」の表示が便利である。この画面では、KWIC コンコーダンスを実行した後に、その用語の前後にどのような語が多く出現していたかを容易に表示することができる。表 4-8 が実際の画面である。「左５」から「右５」まで数値が記入されている。例えば、employment という用語を中心に考え得ると、「左１」に full が 329 回、「右１」に policy が 56 回、それぞれ出現して

[86] Keyword in context の略である（石田・金編 2012: 214）。

いることがわかる。unemployment を軸とすると、「右１」に rate が 58 回、insurance が 34 回、「左１」に mass が 33 回、出現している。industry ならば、other industry, particular industry, location [of] industry の組み合わせが多くなる。demand の場合は、demand [for] labour, effective demand, overseas demand などの組み合わせが多くなる。policy の場合は、full [employment] policy, economic policy の組み合わせが多い。labour の場合は、labour market, demand (supply) [for, of] labour, labour force の組み合わせが多い。outlay の場合は、public (private, total) outlay の組が多い。

	抽出語	品詞	合計	左合計	右合計	左5	左４	左３	左２	左１	右１	右２	右３	右４	右５	スコア
1	full	Adj	338	333	5	0	2	0	2	329	0	1	1	0	3	331.933
2	policy	Noun	106	45	61	5	3	37	0	0	56	0	2	1	2	71.4
3	exchange	Noun	23	0	23	0	0	0	0	0	23	0	0	0	0	23
4	condition	Noun	32	29	3	3	0	24	2	0	0	1	0	1	1	10.55
5	outlay	Noun	27	13	14	0	3	9	1	0	1	1	3	5	4	8.8
6	volume	Noun	11	5	6	0	0	0	5	0	6	0	0	0	0	8.5
7	rate	Noun	12	5	7	0	0	2	3	0	5	0	1	1	0	7.75
8	problem	Noun	18	10	8	0	1	6	3	0	2	0	2	4	0	7.417
9	unemployment	Noun	21	7	14	0	2	5	0	0	0	7	1	2	4	7.3
10	casual	Adj	8	7	1	0	0	0	0	7	0	0	0	0	1	7.2
11	level	Noun	14	12	2	0	0	0	12	0	1	0	0	0	1	7.2
12	war	Noun	23	7	16	4	1	2	0	0	0	5	2	2	7	6.783
13	productive	Adj	6	6	0	0	0	0	0	6	0	0	0	0	0	6
14	change	Noun	14	10	4	0	1	1	8	0	0	0	0	2	2	5.483
15	industry	Noun	16	7	9	0	1	3	3	0	0	0	5	3	1	5.367
16	industrial	Adj	9	4	5	0	0	1	0	3	0	3	0	2	0	5.333
17	peace	Noun	12	1	11	1	0	0	0	0	0	9	1	0	1	5.233
18	number	Noun	14	12	2	3	2	1	6	0	0	0	2	0	0	5.1
19	home	Noun	11	1	10	1	0	0	0	0	0	7	3	0	0	4.7
20	country	Noun	17	10	7	3	2	5	0	0	0	0	3	1	3	4.617

employment | unemployment | industry | demand | war | policy | labor | country

表 4-8　コロケーション分析（employment の場合）

4-6　複合語（専門用語）の抽出

この項では KH Coder に組み込まれているが、別のソフト・モジュールを用いた分析を例示する。専門用語（キーワード）自動抽出システム「TermExtract」[87]である。「前処理」→「複合語の検出」で実行できる。

このシステムの出発点は、専門用語[88]の多くが複合語、とりわけ複合名詞である点に注目したことである。2-2 で前述したように、ある理念・概念・思想を最も象徴的に示すのは名詞であることが多い。そして思想の担い手は、最小単位の名詞を単独で頻繁に用いるだけでなく、その名詞を前後に含む新しい複合名詞を用いることも多い（中川・湯本・森 2003: 27）。本稿の例で記述すれば、「雇用」から派生した「完全雇用」「不完全雇用均衡」、「流動性」から派生した「流動性選好」「流動性選好関数」などである。そこでこうした複合名詞を専門用語として組み立て、対象テキストから重要度を添えて自動的に検出できないかという発想が出てくる。

今までの形態素解析では言葉の最小単位まで切り取るので、単語を組み合わせた複合語には直接対応していなかった。また、専門的な文章を解析するためには、どの用語が重要であるかを判断する仕組み[89]も必要となる。こうした要望に応えるため、「専門用語（キーワード）自動抽出システム」では、次の三段階を取り入れている（中川・湯本・森 2003: 28）。第１に、形態素解析による単語の分割（ここまでは通常のテキストマイニングソフトと同一）。第２に、複合語の形成。第３に、テキストにおける複合語の重要度計算。第２段階では、ある最小単位の名詞にどれほど多くの別の最小単位の名詞が連結するかという種類数を測り（単語の長さに依存させないため、相乗平均を取っている）、さらにその頻度も計算する（連接の種類数と頻度の計算）。第３段階では、この

87 元のシステムは中川裕志（東京大学）・森辰則（横浜国立大学）の両者によって開発された。その後、改良版が出ている。http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html（2017.8.18 閲覧）。

88 専門用語の重要な性質として、①ターム性 termhood : ある表現が対象分野固有の概念とどれだけ関連性を持っているか（出現頻度が主な計算）、②ユニット性 unithood : ある言語単位が対象分野のコーパスの中でどれだけ安定性を持っているか（各言語単位の連接頻度が主な計算）がある（Kageura & Umino 1996: 282-283）。石橋ほか（2014: 1）はさらに、論文の題名と概要しか利用できない場合は、③コミュニティ性 : 専門家集団と一般集団で、ある専門用語の出現分布パターンが異なること、が有力であると主張している。

89 久保・辻・杉本（2010: 17）によれば、専門用語抽出は、①品詞パターンに基づく手法、②語構成要素の連結に注目した手法、③語長と単一分野のコーパスにおける頻度に基づく手法、④複数分野のコーパスにおける頻度に基づく手法、の４つに分かれる。TermExtract は②に相当する。

ように形成された複合語がテキストの中でどれほどの頻度で出現するかを加味して、重要度を判定する。このような複合語の抽出によって、「単純に頻度が高いだけの名詞」の「スコアを低くする効果がある点が有利」（中川・湯本・森 2003: 40）となる。

このモジュールを用いて『自由社会』を分析したところ、次の結果をみた。スコアとは、その複合語の重要度を数値化したものである。

	『自由社会』	スコア		『一般理論』	スコア
1	full employment	38283.435		rate of interest	59714.952
2	full employment policy	5163.317		marginal efficiency of capital	6579.732
3	international trade	3992.062		quantity of money	6536.692
4	labor market	3902.355		full employment	5807.993
5	first world war	2373.954		effective demand	4334.557
6	unemployment rate	2373.824		marginal efficiency	3995.97
7	united states	2118.38		efficiency of capital	3858.848
8	private investment	1838.44		rate of investment	3666.585
9	public outlay	1637.109		terms of money	2209.979
10	multilateral trade	1495.281		user cost	2105.101
11	unemployment insurance	1306.352		new investment	2068.328
12	mass unemployment	1276.042		real wage	1747.277
13	other countries	1233.579		real income	1619.279
14	supply of labor	1230.923		classical theory	1612.696
15	total outlay	1059.734		level of employment	1523.044
16	private outlay	1058.47		volume of employment	1511.791
17	cyclical fluctuation	1029.943		net income	1390.755
18	unemployment rates	1021.576		capital equipment	1290.096
19	social insurance	922.537		supply price	1227.051
20	total demand	873.033		aggregate investment	1184.698

表 4-9　複合語検出（上位 20 語）

表 4-9 からは非常に興味深い結果がわかる。次の 6 点にまとめておこう。第 1 に、専門語として認知された頂点には、full employment そして full employment policy がある。これは題材からしても当然である。第 6 位、第 11 位、第 12 位、第 18 位にもそれぞれ雇用（失業）関連の重要語が並んでいることも指摘しておこう。第 2 に、第 3 位の international trade や第 10 位の multilateral trade や第 13 位の other countries を鑑みると、他の国との関係、特にその貿易を十分考慮した上での完全雇用政策という力点が浮かび上がってくる。閉鎖体系とされる『一般理論』とは際立つ特徴である。この解析結果は、ベヴァリッジが完全雇用政策（社会保障政策と並んで福祉国家理念の屋台骨）を、国際貿易が含まれる国際関係の中で考察し、ある一国の内政とは捉えていない、と解釈することができるだろう。

第 3 に、第 4 位と第 5 位に労働市場・労働供給という重要語が入ったように、ベヴァリッジが初期から強調している労働の供給側の問題が――ケインズ的分析を取り入れた本書でも――なお色濃く残っているのである。第 4 に、総需要に関連する用語が第 15 位、第 16 位、第 20 位という具合に 3 つ入っていることである。供給側よりもスコアとしては下位にあるが、この事実を重要度の差と解釈すべきかどうかは不明瞭である。第 5 に、その他として、第一次世界大戦、アメリカ合衆国という用語、さらに景気循環や社会保険という用語が重要と判定されていることである。特に、後者は景気循環では対応できない部分は社会保険によって失業という悪弊の緩和を目指す本書の内容と合致している。

第 6 に、日本語では「完全雇用」「国際貿易」「社会保険」「総需要」のように名詞の連結となるが、英語では同じ専門用語が「形容詞＋名詞」と結びつくパターンが多くなっている（例外は「失業保険」「循環的変動」などである）。この例は、専門用語の重要度を測る際に、英語では「名詞＋名詞」のみの考察では不十分であることを示唆している。

複合語の検出で際立つのは、むしろ『一般理論』の重要語である。『自由社会』と複合語として重なる用語は、上位 20 位では full employment という 1 つしかないが、雇用・投資・総需要などに関する用語は、（別表現とはいえ）重なっている。しかし、最も重要なのは『一般理論』の最重要複合語として、rate of interest[90]があり、その次に marginal efficiency of capital, quantity of money

90 この複合語のスコアが非常に高いのは、別の理由がある――例えば、rate, of, interest

が続くという事実である。つまり、投資の決定要因として、利子率と資本の限界効率によって投資が定まる（さらにその背後には貨幣供給量がある）という因果が読み取れるのである。その他、使用者費用・実質賃金・古典派経済学など、『一般理論』では馴染みの用語が上位に入っている。足りないのは貨幣需要を決定づける「流動性選好」、投資と所得を関係づける「乗数」という用語であろう[91]。このようなスコア判定から、雇用を決定づける貨幣的な要因という理論的な関心が『一般理論』では前面に出ている、と評価することも可能であろう。ケインズ経済学の本質が乗数にあるのか、流動性選好にあるのかという問題設定が伝統的に続いていた（Minoguchi 1988）。複合語の重要度という解析からは、利子率の決定的重要性が指摘されたことになる。それに対して、『自由社会』では雇用を決定する深層の要因に関する用語は、重要語としては上位に来ていない。

1	employment policy	5811.61
2	labour market	3435.65
3	international trade	2647.27
4	unemployment rate	2408.01
5	private investment	1361.83
6	supply of labour	1190.22
7	unemployment insurance	1190.14
8	total outlay	1140.09
9	private citizens	1045.27
10	unemployment rates	1043.66
11	cyclical fluctuation	1029.16
12	private enterprise	972.65
13	public outlay	914.66
14	mass unemployment	892.76
15	rate of unemployment	732.87
16	total demand	720.15
17	effective demand	701.57

の結びつきが非常に強く、他の組み合わせでは使われない――かもしれない。

[91] 理論の柱である両者がこの表に入っていない理由は不明である。

18	employment exchanges	581.91
19	economic system	581.35
20	cent of unemployment	537.53

表 4-10　複合語検出（ウェブ版、上位 20 語）

同じサイトにある簡易版[92]を用いて、同じテキストを解析してみた。これはモジュールをパソコンにインストールすることなく、ウェブ上ですべての解析を簡易に行うものである。その結果は複合語のみを取り出さず、単独の名詞を含んだ形で、重要スコア順に並べられていた。employment が 46736.93 スコア、unemployment が 29461.14、labour が 14080.16、demand が 10134.00、war が 9662.60、policy が 7207.31、Britain が 6115.41 となっている。この解析表から複合語のみを取り出したのが表 4-10 である。表 4-9 と比べてみると、full という形容詞が脱落し、いくつかの順位は入れ替わり（例：supply of labour は表 4-9 では第 14 位、表 4-10 では第 6 位）、いくつかは新しい単語（例：第 9 位の private citizens）が入っているものの、おおむね分析結果は変わらないようだ。テキストマイニングの解析を行う際は、１つの方法だけに偏らず、代替可能ないくつかの計算方法を試してみて、その最大公約数とも言えそうな妥当な解釈を施す必要があると言えるだろう。特に、順位（ランキング）に過剰な信頼を置くことは危険性を伴う。

第５節　結論と今後の課題

本節は以上の議論をまとめ、今後の課題を付したうえで、結論を述べる。

5-1　考察のまとめ

本稿では主に、次の６つ方法でテキストを解析した。それぞれの方法を一般的に略記して、テキストマイニングに応用する際の留意点を述べる。

【１】頻度分析：各品詞別に、ある単語がどの程度、頻出しているかを単純に数え上げる分析である。特に、名詞・固有名詞・形容詞に注目するだけでも、十分に大きな発見になりうる（動詞は除外してある）。ただし、最後の複合語の抽出と組み合わせた時に、より効果を発揮する。

[92] http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html（2017.8.18 閲覧）。

【２】階層的クラスター分析：共起（ある単語とある単語が同時に出現していること）を重視する係数を用いた類似度によって、テキスト全体を機械的にグループ化する分析である。図が大きく複雑で、様々な方法があり、品詞やクラスター数も選択できるため、結果が大きく変わる。グループ化の発想を豊かにするという程度で留めておく。

【３】共起ネットワーク分析：単語と単語（または見出し語）との共起関係を視覚化する分析である。様々な中心性があるので留意すべきだが、円の大きさや太さ、単語の位置関係など、視覚化として最も優れている。

【４】特徴語分析：Jaccard 係数を用いた類似性を計算して、テキストの単元ごとの特徴を上位から抽出する分析である。章や部という部分を持つテキスト（原典）にとって、全体とその単元の相対的な特徴を掴むのに便利である。

【５】コロケーション統計：特徴語を特定化した上で、その前後にどのような単語が頻出しているかを一覧にする分析である。特徴語と結びつく単語を特定化して、しかも文脈に戻ることができるため、テキスト解析とテキスト原典の往復に便利な機能である。

【６】複合語の抽出：最小単位の単語の接合具合から複合語を専門用語として特定化し、重要度を付して抽出する分析である。単なる頻度分析を越えて、キーワードとしてテキストから抽出できることが最大の利点である。

『自由社会』を『一般理論』との関係で解析した結果を、上記の６つに対応して要約しておこう。

【１】頻度分析：『自由社会』では employment が１位、unemployment が２位、industry が３位、demand が４位、war が５位という頻度順位（名詞）を持つ。social, national, economic, international が形容詞としては上位にくる。『一般理論』とは employment, demand, labor, investment, rate の５つが上位 20 で重なっている。『一般理論』で頻出する上位 20 のうち、interest, money,

capital, income, level, consumption の 6 個はその書評では見逃していないが、『自由社会』では脱落している。『一般理論』では marginal, aggregate, effective, classical という極めて理論的な専門用語を形成する形容詞が上位に来ている。固有名詞は人名のみを考え、6 名を抽出した。

以上より、理論指向の『一般理論』と比して、その貨幣概念と強く関わる単語は『自由社会』では抜け落ち、投資以外の総需要項目は outlay という単語で置き換わっていると判断できる（仮説の要素１）。また、『自由社会』では socail(ize)に関係する単語も頻出する（仮説の要素２）。

【２】階層的クラスター分析：クラスターを 10 にすると、一番近い単語は full と employment、二番目は labor と demand（さらには両者と industry）、三番目は private と public（さらには両者と investment）、四番目が international と trade（さらには両者と country）、五番目が war と peace であることが図から判明した。最初の３つは『自由社会』の題材そのものから容易に類推されるが、四番目と五番目はテキストマイニングによる発見とも見なせる。すなわち、ベヴァリッジが戦争と平和の移行期において、国際貿易を重視しながら、一国の雇用政策を考えていたという解釈である（仮説の要素２）。

【３】共起ネットワーク分析：『自由社会』で中心性を持つ単語は、employment, unemployment, private の３つである。特に３番目は「完全雇用政策」という用語にも近いと見なされ、１つの解釈として、この政策は民間とも大いに関係があるというメッセージと捉えられる（仮説の要素２）。対照的に、『一般理論』で中心性のある単語は employment しかなく、雇用がどのように決まるかという理論的関心が前面に出ている。サブグラフ検出による『自由社会』の類語ブロックも、政策指向を裏付ける。

【４】特徴語分析＋【５】コロケーション統計：微少な数値（Jaccard 係数）の差をどのように解釈するかという問題は残るが、employment, unemployment, industry, demand, war, policy, labor, country の６つを実際に特徴語として指定して、その前後にどのような単語が結びつくかを実証した。unemployment insurance, location of industry, labor force など、ベヴァリッジが初期から持っていた社会保障との結びつきや労働の供給側の考察が窺える（仮説の要素２）。

【６】複合語の抽出：上位にくる専門用語（複合語）を鑑みて、ベヴァリッジが完全雇用政策――社会保障政策と並んで福祉国家理念の屋台骨――を、国際貿易が含まれる国際関係の中で考察し、ある一国の内政とは捉えていない、と解釈できる（仮説の要素２）。ベヴァリッジが初期から強調している労働の供給側の問題が――ケインズ的分析を取り入れた本書でも――なお色濃く残っている（仮説の要素２）。『一般理論』では、雇用を決定づける貨幣的な要因という理論的な関心が『一般理論』では前面に出ている、と評価する。利子率が特に重要である（仮説の要素１）。

以上から、仮説の要素１――失業の本質（貨幣経済）に注目するよりは、有効需要という表面的な原因に着目する――も、要素２――ベヴァリッジ自身の社会観と処方箋を依然として保持している――も、十分に検証できたと結論する。

全体として、テキストマイニングという量的方法と従来の経済学史の方法とをうまく接合させることが可能であった。前者による解析結果は、小峯（2007）の見解――例：『自由社会』をケインズ主義の全面的受容と見たり、逆に独自性がないとしたりした先行研究を否定――を強化するだけでなく、新たに次の２点も発見された。１つは、ベヴァリッジが平和～戦争という移行期にあって、国際関係を強く睨みながら完全雇用政策を唱道していたことである。もう１つは、比較対象とした『一般理論』の構造が明瞭となったことである。つまり、雇用決定の背後にある貨幣的要因を示す専門用語に重みが与えられ、その部分が脱落した『自由社会』の政策指向が浮き彫りになったのである。

5-2　課題と方向性

最後に、技術的な部分と研究対象の部分に分けて、それぞれ今後の課題と方向性を示唆しておこう。

テキストマイニングの技術的な側面で、本研究にはまだ課題も多い。７つ挙げておこう。

第１に、那須川（2006: 45）も注意喚起しているように、テキストマイニングの弱点[93]を意識した上で、分析目的を明確に設定し、文書の適切な範疇分けによってデータを構造化することが必須である。第２に、テキストの前処理自体に

[93] 頻度が少ないが重要な用語が分析から抜け落ちること、同義語や否定語の扱いに難点があること、言語化されていない暗黙の意図をつかめないこと、などがある（末田ほか編 2011: 239-240; 田崎編 2015: 200）。

非常に時間がかかり、しかもOCRから整形したテキストでもまだ誤植が発見されるなどの問題が残っている。これらを解決するにはある程度の人海戦術が必要となるだろう。第3に、ノイズを取り除く作業もつぶさに記録しておく必要がある。なぜなら、再現性を確保し、同時に研究不正を防ぐためにも、作業手順の記録は不可欠だからである。

第4に、今回は行わなかった構文解析も、形態素解析に加えて視野に入れる必要がある。構文解析 syntactic analysis は句（英語）や文節（日本語）を単位として分析し、日本語の場合は係り受け――文節と文節の結びつき――などの関係から文の構造を調べる（金 2009: 33; 石田・金編 2012: 130）。第5に、テキストの対象を本研究では本文のみとしている。つまり、補遺や付録や注や参考文献を――主に、時間節約という理由から――すべて省いた形で分析を出発している。原典研究には注や参考文献も不可欠であり、少なくとも注を入れた場合と入れない場合の比較研究は必要となるだろう。第6に、本稿で用いた技術的な方法の操作や選択が適切かどうか、常に検証していく必要があるだろう。完成されたソフトやモジュールに頼るのは時間の節約になるが、本稿は初期の段階として、なぜそのような計算を行うのかという原理を少しでも理解するため、基本的な概念や操作をやや詳しく記述した。それでもまだ誤解や不十分さは残っているだろう。第7に、本稿で挙げた以外の技法も試す必要があるかもしれない。例えば、あるアイデアが科学者集団に伝播していく過程を捉える指標として、《共引用度》[94]も注目されている。その際に、《引用》の知識社会学的な意味づけ[95]も必要となるだろう。また、対象がテキストというよりは人間関係の場合、社会ネットワーク分析も有用かもしれない。

最後に、テキストマイニングと組み合わせた研究対象の部分は、次の3つに応用が考えられる。

第1に、4-1で述べたように、完全雇用という戦後計画は、労働党・政府・ベヴァリッジ私的委員会の三者が争うように1944年に公表した。『一般理論』を基準とした時に、それぞれどのような特徴が発見できるだろうか。第2に、完

---

94 ある論文で同時に引用されている論文の数（Nab）を、それぞれの論文の被引用数（Naと Nb）の積の平方根で割った数値。共引用度＝Nab／（Na×Nb）$^{1/2}$。この数値が0.30を越えていれば、2つの論文は強く関係している。三輪・安藤（2012: 331）と佐々木（2015: 15）を見よ。

95 藤垣（2003: 60, 61, 69）によれば、引用とは、他の論文群の中で当該論文の肯定的・否定的位置づけを示す《方位磁石》compass である。引用によって、知識内容が常に再構成される。この意味で、科学は「堅い一枚岩」なのではなく、常に「作動中」である。

全雇用政策と並ぶもう１つの柱である社会保障政策に関して、『ベヴァリッジ報告』（1942）に対する政府の公式な反応が白書『社会保険』（1944）であった。この白書はどの程度、ベヴァリッジの勧告書から遠ざかったのだろうか。さらにアトリー内閣で国民保険法が 1946 年に成立したが、その法律条文はどのような特徴を持っているのだろうか。Parliamentary Papers[96]の検索も必要になるだろう。第３に、『自由社会における完全雇用』は他人の借り物であるという悪評があるとしたら、その程度を確定することは可能だろうか。本稿では試みていない文体分析[97]が必要となる。

テキストマイニングが内包している量的分析（技術的・統計的部分）は、その適切さ・不適切さを含めて、従来の経済学史研究が持っている質的手法と結合[98]を試みることで、より深みのあるテキストの解釈をもたらす可能性を秘めている。また経済学史の研究者と他分野（統計学、情報処理学、計量言語学など）の研究者を結びつけ、より説得的な、新しい知見をもたらす触媒ともなりうるだろう。ただし、こうした課題を克服したり拡張したりする過程で、テキストマイニングの技術的な難点も改善され、同時に経済学史により適合的な手法もより洗練度を増していかなければならない。

補遺　Word Clouds を用いた視覚化

Word Clouds は用語の出現頻度を視角的に表現した図解であり、様々なフリーソフト・フリーサイトがある。そのうち１つ[99]を使って、『自由社会』と『一般理論』のテキストを可視化しておく。どちらの図が『自由社会』を表すのか、本稿の研究からも明らかであろう。

---

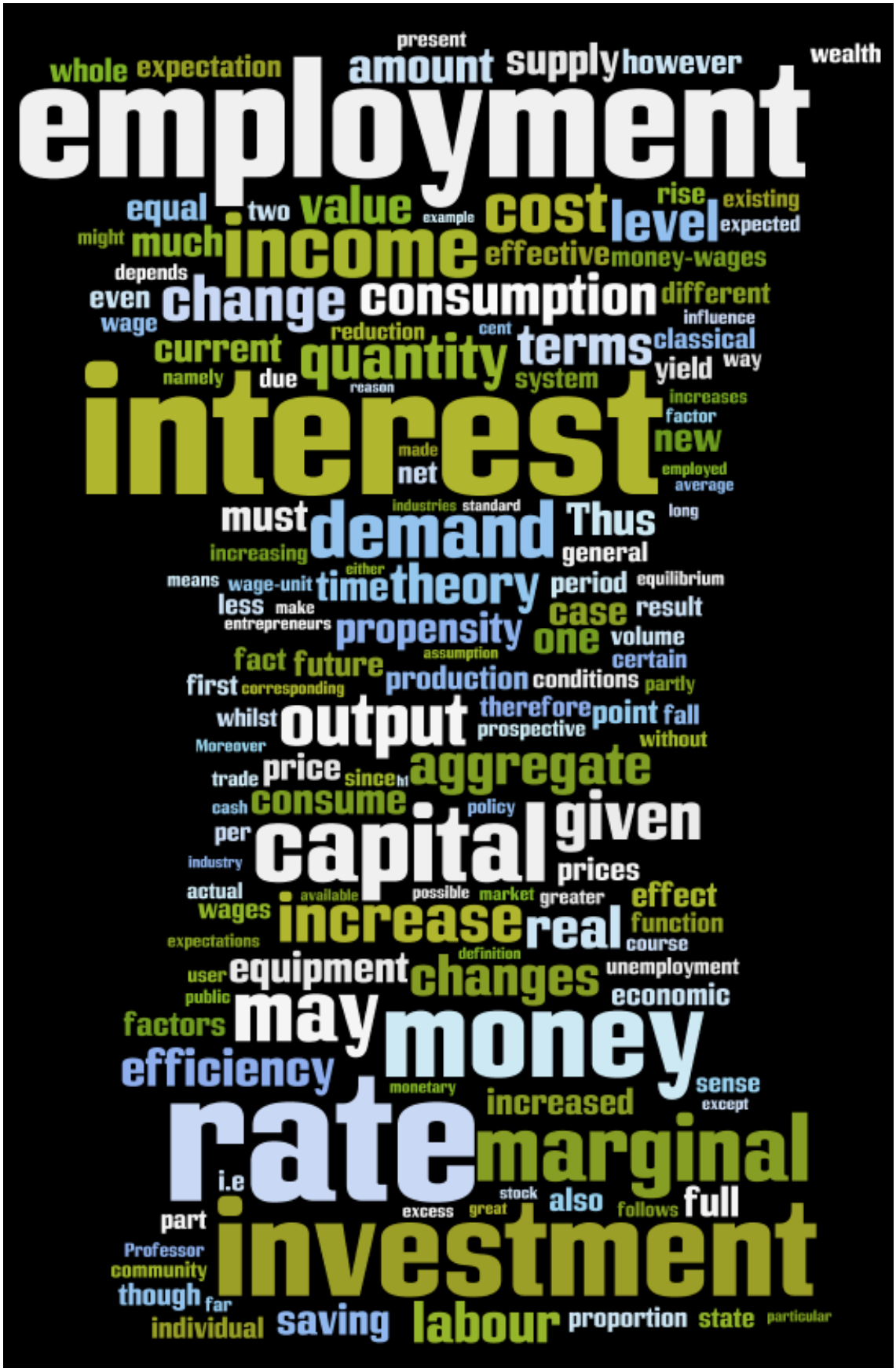
96 例えば、http://hansard.millbanksystems.com/によって、1803 年から 2005 年までの文書が全文検索可能である。

97 単語や文の長さ、品詞の分布、語彙などを比べる。日本語文献では、助詞の使用分布、漢字〜かなの比率、読点の打ち方などが有効である（石田・金編 2012: 57-62）。

98 Creswell (2015: 3, 49, 111／訳 3, 55, 124)によれば、両者を単に並立させるだけでは不十分で、両者を統合する目的とデザインを明瞭にしなければならない。

99 Wordle: http://www.wordle.net/ （2017.8.28 閲覧）


people
still
national
jobs
cent
resources
cyclical
less
general
good
home
expenditure
number
available
first
three
increase
conditions
necessary
spending
might
industry
whole
measures
changes
rather
prices
fluctuation
time
World
War
great
demand
multilateral
order
facts
mean
system
different
figures
possible
unemployment
Part
many
another
consumption
United
citizens
international
industrial
interest
two
taxation
Government
States
shown
depends
even
war
may
made
wages
place
second
wars
much
total
country
far
public
community
free
services
need
problem
investment
goods
taken
particular
unemployed
outlay
make
set
full
Britain
market
British
use
peace
must
though
means
also
main
high
years
year
labour
important
work
past
power
given
industries
change
rate
rates
insurance
income
production
period
business
Report
countries
trade
way
men
per
nearly
control
private
part
large
without
service
every
exports
done
money
employment
world
persons
essential
social
needs
depression
one
State
economy
greater
economic
policy
making
Table
local
supply
imports
take
millions
new


present
whole
expectation
amount
supply
however
wealth
employment
equal
two
value
example
cost
rise
existing
level
expected
might
much
income
effective
money-wages
depends
even
change
consumption
different
influence
wage
reduction
cent
current
quantity
terms
classical
way
namely
due
reason
system
yield
interest
increases
factor
new
made
net
employed
average
industries
standard
long
must
demand
Thus
increasing
general
either
means
wage-unit
time
theory
period
equilibrium
less
make
entrepreneurs
propensity
case
result
one
volume
assumption
certain
fact
future
production
conditions
partly
first
corresponding
whilst
output
therefore
point
fall
prospective
Moreover
without
trade
price
since
hl
aggregate
cash
consume
policy
per
capital
given
industry
prices
actual
available
possible
market
greater
effect
wages
increase
real
function
expectations
course
definition
user
equipment
changes
unemployment
public
economic
factors
may
money
efficiency
monetary
sense
except
increased
rate
marginal
i.e
stock
also
follows
full
excess
great
part
investment
Professor
community
though
far
individual
saving
labour
proportion
state
particular

参考文献

Amable, B. (2003) *The Diversity of Modern Capitalism*, Oxford: Oxford University Press. （山田鋭夫ほか訳『五つの資本主義～グローバリズム時代における社会経済システムの多様性』藤原書店、2005 年。）

Beveridge, W. H. (1942) *Social Insurance and Allied Services*, Cmd. 6404, London: His Majesty's Stationary Office. （一圓光彌監訳『ベヴァリッジ報告～社会保険および関連サービス』法律文化社、2014 年。）

Beveridge, W. H. (1944) *Full Employment in a Free Society*, London: George Allen & Unwin Ltd.

Binder, J. M. and C. Jennings (2016) "'A Scientifical View of the Whole': Adam Smith, Indexing, and Technologies of Abstraction", *Journal of English Literary History*, 83(1): 157-180; DOI: 10.1353/elh.2016.0001.

Booth, A. (1989) *British Economic Policy, 1931-49: Was There a Keynesian Revolution?* Hemel Hempstead, Hertfordshire, U. K.: Harvester Wheatsheaf.

Brady, H. E. and D. Collier eds. (2010) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, second edition (first published in 2004), Maryland, U. S.: Rowman & Littlefield Publishers, Inc. （泉川泰博・宮下明聡訳『社会科学の方法論争～多様な分析道具と共通の基準』勁草書房、2008 年、初版の訳）。

Claveau, F. and Y. Gingras (2016) "Macrodynamics of Economics: A Bibliometric History", *History of Political Economy*, 48(4): 551-592, December 2016; DOI: 10.1215/00182702-3687259.

Coats, A.W. ed. (1981) *Economists in Government: An International Comparative Study*, Durham, U.S.: Duke University Press.

Coats, A.W. (1993) *The Sociology and Professsionalization of Economics*, London and New York: Routledge.

Creswell, J. W. (2015) *A Concise Introduction to Mixed Methods Research*, Los Angeles, U. S.: Sage. （抱井尚子訳『早わかり混合研究法』ナカニシヤ出版、2017 年。）

Dalton, H. (1957) *Hugh Dalton Memoirs 1931-1945: The Fateful Years*, London: Frederick Muller Ltd.

Flick, U. (2007) *Designing Qualitative Research*, Los Angeles: SAGE. （鈴木聡志訳『質的研究のデザイン』新曜社、2016 年。）

Fujita, N. (2014) "Historical Evolution of Welfare Policy Ideas: the Scandinavian Perspective", in H. Magara ed. (2014) *Economic Crises and Policy Regimes: The Dynamics of Policy Innovation and Paradigmatic Change*, Cheltenham, U.K.: Edward Elgar: 337-356.

Furner, M.O. and B. Supple eds. (1990) *The State and Economic Knowledge: The American and British Experiences*, Cambridge: Cambridge University Press.

Goertz, G. and J. Mahoney (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton, U. S.: Princeton University Press. （西川賢・今井真士訳『社会科学のパラダイム論争～２つの文化の物語』勁草書房、2015 年）。

Hayek, F. A. (1994) *Hayek on Hayek: An Autobiographical Dialogue,* edited by S. Kresge, & L. Wenar, London: Routledge. （嶋津格訳『ハイエク、ハイエクを語る』名古屋大学出版会、2000 年。）

Harris, J. (1997) *William Beveridge: A Biography*, revised paperback edition, Oxford: Oxford University Press.

Ignatow, G. and R. Mihalcea (2017) *Text Mining: A Guidebook for the Social Sciences*, California, U. S.: SAGE Publications, Inc.

Kageura, K. and B. Umino (1996) "Methods of Automatic Term Recognition: A Review", *Terminology*, 3(2): 259-289; DOI: https://doi.org/10.1075/term.3.2.03kag.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan. （塩野谷佑一訳『雇用・利子および貨幣の一般理論』（ケインズ全集第７巻）東洋経済新報社、1983 年。）

Keynes, J. M. (1973) *The General Theory and After: Part Ⅰ, Preparation*, paperback, the Collected Writings of John Maynard Keynes, vol. XIII, London: Macmillan.

Keynes, J. M. (1979) *Activities 1939-1945: Internal War Finance*, the Collected Writings of John Maynard Keynes, vol. XXII, London: Macmillan.

Keynes, J. M. (1980) *Activities 1940-1946: Shaping the Post-War World: Employment and Commodities*, the Collected Writings of John Maynard Keynes, vol. XXII, London: Macmillan. （平井俊顕・立脇和夫訳『戦後世界の形成　雇用と商品～1940～46 年の諸活動』（ケインズ全集第 27 巻）東洋経済新報社、1996 年）。

Matsuyama, N. (2016) "A Study of Text mining for Research into the History of Economic Thought: The case of Alfred Marshall's *Principles of Economics* (1890)", Discussion Paper, University of Hyogo, No.89, 2016.

Minoguchi, T. (1994) "Some Questions about IS-LM Interpretations of The General Theory", in J.C. Wood ed. *John Maynard Keynes, Critical Assessment*, Second Series, vol. V, London and New York: Routledge.

O'Neill, H. (2016) "John Stuart Mill and the London Library: A Victorian Book Legacy Revealed", *Book History*, 19: 256-283; DOI: 10.1353/bh.2016.0007

Peden, G. C. (1979) *British Rearmament and the Treasury: 1932-1939,* Edinburgh: Scottish Academic Press.

Robbins, L. (1971) *Autobiography of an Economist*, London: Macmillan. （田中秀夫監訳『一経済学者の自伝』ミネルヴァ書房、2009 年。）

Schumpeter, J. A. (1994 [1954]) *History of Economic Analysis*, with an Introduction by Mark Perlman, London: Routledge. （東畑精一・福岡正夫訳『経済分析の歴史』（上）岩波書店、2005 年。）

Scott, J. (2013) *Social Network Analysis*, third edition, first published in 1991, London: SAGE Publications Ltd.

Toye, R. (2003) *The Labour Party and the Planned Economy, 1931-1951*, Woodbridge, Suffolk, U. K.: The Boydell Press.

Wiedemann, G. (2016) *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*, Wiesbaden, Germany: Springer.

Wright, C. (2016) "The 1920s Viennese Intellectual Community as a Center for Ideas Exchange: A Network Analysis", *History of Political Economy*, 48(4): 593-634; DOI: 10.1215/00182702-3687271.

池尾愛子（2006）『日本の経済学〜20 世紀における国際化の歴史』名古屋大学出版会。
石田淳（2017）『集合論による社会的カテゴリー論の展開〜ブール代数と質的比較分析の応用』勁草書房。
石田基広（2008）『R によるテキストマイニング入門』森北出版。
石田基広（2017）『R によるテキストマイニング入門　第２版』森北出版。
石田基広・金明哲編（2012）『コーパスとテキストマイニング』共立出版。
石田基広・小林雄一郎（2013）『R で学ぶ日本語テキストマイニング』ひつじ書房。
石橋和樹・南出直樹・風間一洋・篠田孝祐（2014）「単語のコミュニティ性に基づいた専門用語の抽出」、『人工知能学会全国大会論文集』28 号: 1-4。
岡嶋裕史（2006）『数式を使わないデータマイニング入門〜隠れた法則を発見する』光文社新書。
岡﨑直観（2016）「単語の意味をコンピュータに教える」、『岩波データサイエンス』（岩波書店）、Vol.2: 47-61。
小木しのぶ（2015）「テキストマイニングの技術と動向」、『計算機統計学』28(1): 31-40、2015。
奥野陽・グラム-ニュービッグ・萩原正人（2016）『自然言語処理の基本と技術』翔泳社。
喜田昌樹（2007）『組織確信の認知的研究〜認知変化・知識の可視化と組織科学へのテキストマイニングの導入』白桃書房。
喜田昌樹（2008）『テキストマイニング入門』白桃書房。
喜田昌樹（2010）『ビジネス・データマイニング入門』白桃書房。
金明哲（2009）『テキストデータの統計科学入門』岩波書店。
久保順子・辻慶太・杉本重雄（2010）「異なる学問分野のコーパスを利用した専門用語抽出手法の提案」『情報知識学会誌』30(1): 15-31, 2010-02。
小林雄一郎（2017a）『R によるやさしいテキストマイニング』オーム社。
小林雄一郎（2017b）『仕事に使えるクチコミ分析〜テキストマイニングと統計学をマーケティングに活用する』技術評論社。
小峯敦（2007）『ベヴァリッジの経済思想〜ケインズたちとの交流』昭和堂。

小峯敦（2014）「『ベヴァリッジ報告』(1942)と『雇用政策』白書(1944)：戦後構想（社会保障と完全雇用）における経済助言活動の役割」、『龍谷大学経済学論集』53(1/2): 37-98, 2014-02。
小室達章（2016）「テキストマイニングを活用したリスク概念の分析」、『金城学院大学論集. 社会科学編』12(2): 20-36、2016-03。
小室正紀（2012）「福澤諭吉の経済論における「官民調和」」『日本経済思想史研究』12: 47-57。
菰田文男・那須川哲哉編（2014）『ビックデータを活かす技術戦略としてのテキストマイニング』中央経済社。
佐々木英和（2012）「現代日本語「自己実現」の使われ方に関する基盤的考察～テキストマイニングの手法を用いた質的データの分析」『宇都宮大学教育学部紀要　第 1 部』62 号: 225-238、2012-03。
佐々木英和（2015）「自己実現言説における「社会」の位置づけに関する一考察～テキストマイニング手法による新事実の発見と実証の試み」、『宇都宮大学教育学部研究紀要. 第 1 部』65 号: 229-248、2015-03。
佐々木英和（2016）「自己実現言説における「社会」の意味合いについての歴史的考察～テキストマイニング手法による量的研究と質的研究との接合の試み」、『宇都宮大学教育学部研究紀要. 第 1 部』66 号: 223-248、2016-03。
佐々木基裕（2015）「日本の教育哲学学界におけるポストモダニズム受容～『教育哲学研究』を事例に」『教育・社会・文化：研究紀要』（京都大学大学院教育学研究科　教育社会学講座）15 号: 1-18。
佐藤郁哉（2008）『質的データ分析法』新陽社。
塩野谷祐一（2009）『経済哲学原理～解釈学的接近』東京大学出版会。
鈴木努（2017）『ネットワーク分析　第２版（R で学ぶデータサイエンス８）』共立出版。
下平裕之・小峯敦・松山直樹（2012）「経済学史研究におけるテキストマイニング分析の導入：ケインズ『一般理論』と書評の関係」Discussion Paper Series, Research Group of Economics and Management, No. 2012-E02,Yamagata University（本稿の一部に改訂して収録）。
下平裕之・福田進治（2014）「古典派経済学の普及過程に関するテキストマイニング分析～リカード、ミル、マーティノーを中心に」、『人文社会論叢　社会科学篇』（弘前大学人文学部）31 号: 51-66。

末田清子・抱井尚子・田崎勝也・猿橋順子編（2011）『コミュニケーション研究法』ナカニシヤ出版。
高史明（2015）『レイシズムを解剖する〜在日コリアンへの偏見とインターネット』勁草書房。
田崎勝也編（2015）『コミュニケーション研究のデータ解析』ナカニシヤ出版。
田中京子（2014）「KH Coder と R を用いたネットワーク分析」、『久留米大学コンピュータジャーナル』（久留米大学情報教育センター）28 号: 37-52, 2014-03。
（独）情報処理推進機構編（2017）「71　ビッグデータ分析技術を応用したソフトウェア不具合の分析実施実例」、『先進的な設計・検証技術の適用事例報告書　2016 年度版』：1-16。
http://www.ipa.go.jp/sec/reports/20170302.html（2017.8.15 閲覧）
豊田秀樹編（2008）『データマイニング入門〜R で学ぶ最新データ解析』東京書籍。
中川裕志・湯本紘彰・森辰則（2003）「出現頻度と連接頻度に基づく専門用語抽出」『自然言語処理』10(1): 27-45, 2003-01。
西沢保・服部正治・栗田啓子編（1999）『経済政策思想史』有斐閣。
那須川哲哉（2006）『テキストマイニングを使う技術／作る技術〜基礎技術と適用事例から導く本質と活用法』東京電機大学出版局。
野村康（2017）『社会科学の考え方〜認識論、リサーチデザイン、手法』名古屋大学出版会。
樋口耕一（2014）『社会調査のための計量テキスト分析〜内容分析の継承と発展を目指して』ナカニシヤ出版。
藤垣裕子（2003）『専門知と公共性〜科学技術社会論の構築へ向けて』東京大学出版会。
古谷豊（2014）「テキストマイニングを用いたスミス『国富論』普及の分析」TERG Discussion Papers（東北大学大学院経済学研究科）、No. 325、2014.11。
松尾豊・友部博教・橋田浩一・中島秀之・石塚満（2005）「Web 上の情報からの人間関係ネットワークの抽出」、『人工知能学会論文誌』20(1)E: 46-56。
松村真宏・三浦麻子（2009）『人文・社会科学のためのテキストマイニング』誠信書房。

美馬秀樹・丹治信・増田勝也・太田晋（2012）「近代文献のデジタルアーカイブとテキストマイニング～岩波書店『思想』を題材に」『情報処理学会研究報告　人文科学とコンピュータ研究会報告』2012-CH-95(4): 1-8。
http://www.bookpark.ne.jp/ipsj/
三輪唆矢佳・安藤聡子（2012）「先端研究領域を見いだす「リサーチフロント」分析～ビブリオメトリックスの一事例」『情報管理』(国立研究開発法人　科学技術振興機構）、55(5): 329-338。
村井源編（2014）『量から質に迫る～人間の複雑な感性をいかに「計る」か』新曜社。
山本真照（2011）「テキストマイニング手法の洗練に向けた知識活用方法に関する研究」、『経済科学論究』（埼玉大学経済学会）8 号: 73-85、2011-04。
涌井良幸・涌井貞美（2011）『多変量解析がわかる』技術評論社。

（株）マクロミル社：http://www.macromill.com/method/index.html
TinyTextMiner βversion: http://mtmr.jp/ttm/
統計ソフト R: http://o-server.main.jp/r/about.html
R Commander, 1.8-x version: http://socserv.mcmaster.ca/jfox/Misc/Rc
KH Coder: http://khc.sourceforge.net/